

STIC-Biotech/ChemLib

90253

From: O'Hara, Eileen
Sent: Friday, March 28, 2003 6:18 PM
To: STIC-Biotech/ChemLib
Subject: sequence search for 09/803,589

SEARCH REQUEST FORM-SEQUENCE

Examiner: Eileen O'Hara (Room 10E18) Art Unit 1646

Serial Number: **09/803,589**

**Novel Genes Encoding Proteins Having Prognostic, Diagnostic, Preventive, Therapeutic and Other Uses
McCarthy et al.**

Date: 3/28/03 Phone: 308-3312 CM1 10A01 mailbox 10D19

Please search the following:

SEQ ID NOS: 5 and 6 (nucleic acid and encoded protein)

also, 15 consecutive nucleotides of SEQ ID NO: 5

also, 10 consecutive amino acids of SEQ ID NO: 6

in **commercial and patent** databases.

Please put results on paper.

Thank you!

~~Eileen B. O'Hara~~
~~CM1 10A01 mailbox 10D19~~
~~703-308-3312~~
~~U.S. Patent and Trademark Office~~
~~Art Unit 1646~~
~~eileen.o'hara@uspto.gov~~

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 4/4
Date Completed: 4/14
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 19:33:58 ; Search time 103 Seconds
(without alignments)
2914.918 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcgccagcaggccagc.....gaggatccctaccctccctggc 979

Scoring table: OLIGO_NUC

Gapop 60.0 , Capext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 15

Total number of hits satisfying chosen parameters: 1276

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.9	467	2	US-08-841-349-18
2	19	1.9	489	1	US-08-663-310-3
3	19	1.9	489	2	US-09-006-491-3
4	19	1.9	489	3	US-09-335-919-3
5	19	1.9	972	4	US-08-980-832-33
6	19	1.9	1261	4	US-08-980-832-30
7	19	1.9	1930	1	US-08-544-900-1
8	19	1.9	1930	5	PCT-US95-07874-1
9	19	1.9	2350	4	US-09-187-478-1
10	19	1.9	2350	4	US-09-292-036-1
11	19	1.9	2886	1	US-08-663-310-7
12	19	1.9	2886	2	US-09-006-491-7
13	19	1.9	2886	3	US-09-335-919-7
14	19	1.9	6854	4	US-09-194-905-7
15	19	1.9	8310	3	US-08-870-126-11
16	19	1.9	8310	4	US-09-445-247-11
C 17	19	1.9	33529	4	US-09-144-085-3
18	18	1.8	497	4	US-09-370-838-279
19	18	1.8	697	4	US-09-227-357-96
20	18	1.8	1230	3	US-09-188-930-25
21	18	1.8	1230	3	US-09-188-930-201
22	18	1.8	1454	2	US-08-713-000-7
23	18	1.8	1454	2	US-08-975-316-7
24	18	1.8	1454	4	US-09-211-710-7
25	18	1.8	1454	4	US-09-615-192A-7
26	18	1.8	1932	4	US-09-518-386B-2
27	18	1.8	2129	1	US-08-426-819A-34

28	18	1.8	2574	2	US-08-677-734A-8	Sequence 8, Appli
29	18	1.8	2574	4	US-09-097-053-8	Sequence 8, Appli
C 30	18	1.8	5515	4	US-09-398-193-98	Sequence 98, Appli
31	18	1.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C 32	18	1.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
33	17	1.7	383	4	US-09-370-838-285	Sequence 285, App
34	17	1.7	420	1	US-08-470-179-112	Sequence 112, App
35	17	1.7	420	1	US-08-470-179-148	Sequence 148, App
36	17	1.7	527	4	US-09-370-838-281	Sequence 281, App
37	17	1.7	530	3	US-08-758-662-4	Sequence 4, Appli
38	17	1.7	613	4	US-09-370-838-238	Sequence 238, App
39	17	1.7	624	2	US-08-713-000-9	Sequence 9, Appli
40	17	1.7	624	2	US-08-975-316-9	Sequence 9, Appli
41	17	1.7	624	4	US-09-211-710-9	Sequence 9, Appli
42	17	1.7	624	4	US-09-615-192A-9	Sequence 45, Appli
43	17	1.7	684	2	US-08-975-316-45	Sequence 45, Appli
44	17	1.7	684	4	US-09-615-192A-45	Sequence 45, Appli
45	17	1.7	810	4	US-08-406-030A-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1

US-08-841-349-18/c
; Sequence 18, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISHRA, LOPA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
; FILE REFERENCE: XX/PO4470USO
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: For all n's in this sequence, n=(a or g or c or t)
US-08-841-349-18

Query Match 1.9%; Score 19; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCAGGCGCAG 19
Db 466 GAATTCGGCAGGCGCAG 448

RESULT 2

US-08-663-310-3
; Sequence 3, Application US/08663310
; Patent No. 5811273
; GENERAL INFORMATION:
; APPLICANT: MISAWA, No. 5811273hiko
; APPLICANT: KONDO, Keiji
; APPLICANT: KAJIWARA, Susumu
; APPLICANT: YOKOYAMA, Akihiko
; TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
; TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
; TITLE OF INVENTION: XANTHOPHYLLS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,491
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/663,310
FILING DATE: 23-SEP-1996
APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-348737
FILING DATE: 27-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/109
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..486
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1..486
US-09-006-491-3

Query Match 1.9%; Score 19; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gap

QY 330 GCGGCGCGAGCGCAGGAG 348
|||||
Db 462 GCGGCGCGAGCGCAGGAG 480

RESULT 4
US-09-335-919-3
Sequence 3, Application US/09335919
Patent No. 6150130
GENERAL INFORMATION:
APPLICANT: MISAWA, No. 61501301hiko
APPLICANT: KONDO, Kei11
APPLICANT: KAJIWARA, Susumu
APPLICANT: YOKOYAMA, Akihiro
TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
TITLE OF INVENTION: XANTHOPHYLLS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,919

```

;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/663,310
;; FILING DATE: 23-SEP-1996
;; APPLICATION NUMBER: WO PCT/JP94/02220
;; FILING DATE: 26-DEC-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 6-235917
;; FILING DATE: 05-SEP-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 5-348737
;; FILING DATE: 27-DEC-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 49441/109
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 489 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..486
;; FEATURE:
;; NAME/KEY: mat_peptide
;; LOCATION: 1..486
US-09-335-919-3

Query Match 1.9%; Score 19; DB 3; Length 489;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGAGCGCAGGAG 348
|||||
DB 462 GCGGGCGGAGCGCAGGAG 480

RESULT 5
US-08-980-832-33
; Sequence 33, Application US/08980832B
; Patent No. 6291204
; GENERAL INFORMATION:
; APPLICANT: Pasamontes, Luis
; TITLE OF INVENTION: Improved Fermentative Carotenoid Production
; FILE REFERENCE: Improved Fermentative Carotenoid
; CURRENT APPLICATION NUMBER: US/08/980,832B
; CURRENT FILING DATE: 1997-12-01
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 972
; TYPE: DNA
; ORGANISM: E-396
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Unknown
US-08-980-832-33

Query Match 1.9%; Score 19; DB 4; Length 972;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGAGCGCAGGAG 348
|||||
DB 882 GCGGGCGGAGCGCAGGAG 900

RESULT 6
US-08-980-832-30
; Sequence 30, Application US/08980832B
; Patent No. 6291204
; GENERAL INFORMATION:
; APPLICANT: Pasamontes, Luis
; TITLE OF INVENTION: Improved Fermentative Carotenoid Production
; FILE REFERENCE: Improved Fermentative Carotenoid
; CURRENT APPLICATION NUMBER: US/08/980,832B
; CURRENT FILING DATE: 1997-12-01
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 1261
; TYPE: DNA
; ORGANISM: Alcalligenes PC-1
US-08-980-832-30

Query Match 1.9%; Score 19; DB 4; Length 1261;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGAGCGCAGGAG 348
|||||
DB 1226 GCGGGCGGAGCGCAGGAG 1244

RESULT 7
US-08-544-900-1
; Sequence 1, Application US/08544900
; Patent No. 5736331
; GENERAL INFORMATION:
; APPLICANT: Lin, Stanley Li
; APPLICANT: Rochofsky, Marnie Lynn
; TITLE OF INVENTION: Method for Identifying Nucleic
; TITLE OF INVENTION: Acids Encoding c-fos Promoter
; TITLE OF INVENTION: Activating Proteins
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; ADDRESSEE: Patent Department K-6-1 (1990)
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,900
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/272,412
; FILING DATE: 8-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Thompson, Paul A.
; REGISTRATION NUMBER: 35,385
; REFERENCE/DOCKET NUMBER: OC0439K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908 298 5150
; TELEFAX: 908 298 5388
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1930 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-544-900-1

Query Match 1.9%; Score 19; DB 1; Length 1930;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 TTTTGGTGTCTGTCTGTCT 125
|||||
DB 1728 TTTTGGTGTCTGTCTGTCT 1746

RESULT 8
PCT-US95-07874-1

; Sequence 1, Application PC/TUS9507874

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Method for Identifying Nucleic Acids Encoding c-fos Promoter A

; NUMBER OF SEQUENCES: 2

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Apple Macintosh

; OPERATING SYSTEM: Macintosh 7.1

; SOFTWARE: Microsoft Word 5.1a

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/07874

; FILING DATE: -June-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/272,412

; FILING DATE: 8-JUL-1994

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1930 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

PCT-US95-07874-1

Query Match 1.9%; Score 19; DB 5; Length 1930;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 TTTTGGTGTCTGTCTGTCT 125
|||||
DB 1728 TTTTGGTGTCTGTCTGTCT 1746

RESULT 9

US-09-187-478-1

; Sequence 1, Application US/09187478

; Patent No. 6348329

; GENERAL INFORMATION:

; APPLICANT: Schmidt, Brian F.

; APPLICANT: Allen, Margaret L.

; TITLE OF INVENTION: Connective Tissue Growth (CTGF) And Methods Of Use

; FILE REFERENCE: 08766/004001

; CURRENT APPLICATION NUMBER: US/09/187,478

; CURRENT FILING DATE: 1998-11-06

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 2350

; TYPE: DNA

; ORGANISM: No. 6348329mal Rate Kidney Fibroblast

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (212)..(1252)

US-09-187-478-1

Query Match 1.9%; Score 19; DB 4; Length 2350;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCAGGAGCCAG 19
|||||
DB 1 GAATTCGGCAGGAGCCAG 19

RESULT 10

US-09-292-036-1

; Sequence 1, Application US/09292036

; Patent No. 6358741

; GENERAL INFORMATION:

; APPLICANT: FIBROGEN, INC

; APPLICANT: SCHMIDT, Brian

; APPLICANT: ALLEN, Margaret

; APPLICANT: SVERDRUP, Fran

; APPLICANT: CARMICHAEL, David

; TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR (CTGF) AND METHODS OF USE

; FILE REFERENCE: FIB100-1

; CURRENT APPLICATION NUMBER: US/09/292,036

; CURRENT FILING DATE: 1999-04-14

; PRIOR APPLICATION NUMBER: US 09/292,036

; PRIOR FILING DATE: 1999-04-14

; PRIOR APPLICATION NUMBER: US 09/187,478

; PRIOR FILING DATE: 1998-11-06

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 2350

; TYPE: DNA

; ORGANISM: Rat

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (212)..(1252)

US-09-292-036-1

Query Match 1.9%; Score 19; DB 4; Length 2350;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCAGGAGCCAG 19
|||||
DB 1 GAATTCGGCAGGAGCCAG 19

RESULT 11

US-08-663-310-7

; Sequence 7, Application US/08663310

; Patent No. 5811273

; GENERAL INFORMATION:

; APPLICANT: MISAWA, No. 5811273hiko

; APPLICANT: KONDO, Keiji

; APPLICANT: KAJIWARA, Susumu

; APPLICANT: YOKOYAMA, Akihiro

; TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF

; TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/663,310

; FILING DATE: 23-SEP-1996

; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA: WO PCT/JP94/02220
;; FILING DATE: 26-DEC-1994
;; PRIOR APPLICATION DATA: JP 6-235917
;; FILING DATE: 05-SEP-1994
;; PRIOR APPLICATION DATA: JP 5-348737
;; FILING DATE: 27-DEC-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 49441/109
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2886 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; US-08-663-310-7

Query Match 1.9%; Score 19; DB 1; Length 2886;
Best Local Similarity 100.0%; Pred. NO. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGCGCGCAGGAG 348
|||||
DB 1325 GCGGGCGGCGCGCAGGAG 1343

RESULT 12
US-09-006-491-7
Sequence 7, Application US/09006491
Patent No. 5972690
GENERAL INFORMATION:
APPLICANT: MISAWA, No. 5972690ihiko
APPLICANT: KONDO, Keiji
APPLICANT: KAJIWARA, Susumu
APPLICANT: YOKOYAMA, Akihiro
TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,491
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/663,310
FILING DATE: 23-SEP-1996
APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-348737
FILING DATE: 27-DEC-1993
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/109
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399

;; FILING DATE: 27-DEC-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 49441/109
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2886 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; US-09-006-491-7

Query Match 1.9%; Score 19; DB 2; Length 2886;
Best Local Similarity 100.0%; Pred. NO. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGGCGGCGCGCAGGAG 348
|||||
DB 1325 GCGGGCGGCGCGCAGGAG 1343

RESULT 13
US-09-335-919-7
Sequence 7, Application US/09335919
Patent No. 6150130
GENERAL INFORMATION:
APPLICANT: MISAWA, No. 6150130ihiko
APPLICANT: KONDO, Keiji
APPLICANT: KAJIWARA, Susumu
APPLICANT: YOKOYAMA, Akihiro
TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
TITLE OF INVENTION: XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,919
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/663,310
FILING DATE: 23-SEP-1996
APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-348737
FILING DATE: 27-DEC-1993
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/109
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399

TELEX: 904136
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2886 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-335-919-7

Query Match 1.9% Score 19; DB 3; Length 2886;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GCGGCCGAGCGCGCAGGAG 348
DB 1325 GCGGCCGAGCGCGCAGGAG 1343

RESULT 14

US-09-194-905-7
Sequence 7, Application US/09194905
Patent No. 6306627
GENERAL INFORMATION:

APPLICANT: DECKER, Heinrich
TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
TITLE OF INVENTION: GLA.O AND THEIR USE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/194,905

FILING DATE: 29-JUL-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/EP97/02826

FILING DATE: 30-MAY-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE 19622783.6

FILING DATE: 07-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Granados, Patricia D.

REGISTRATION NUMBER: 33,683

REFERENCE/DOCKET NUMBER: 026083/0193

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 6854 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-194-905-7

Query Match 1.9% Score 19; DB 4; Length 6854;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 GCGAGCTCGCTCGCTGC 475
DB 2753 GCGAGCTCGCTCGCTGC 2771

RESULT 15

US-08-870-126-11/c

Sequence 11, Application US/08870126

Patent No. 6048702

GENERAL INFORMATION:

APPLICANT: Prendergast, George C.

APPLICANT: Sakamuro, Daitoku

TITLE OF INVENTION: Murine and Human Box-Dependent

TITLE OF INVENTION: MYC-Interacting Protein (Bin1) and Uses Therefor

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Cntr, P O Box 457

CITY: Spring House

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/870,126

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/435,454

FILING DATE: 05-MAY-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/652,972

FILING DATE: 24-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Kodroff, Cathy A.

REGISTRATION NUMBER: 33,980

REFERENCE/DOCKET NUMBER: WST60CUSA

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 8310 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: exon

LOCATION: 680..765

OTHER INFORMATION: /note= "exon 7"

FEATURE:

NAME/KEY: exon

LOCATION: 1052..1127

OTHER INFORMATION: /note= "exon 8"

FEATURE:

NAME/KEY: exon

LOCATION: 2503..2585

OTHER INFORMATION: /note= "exon 9"

FEATURE:

NAME/KEY: exon

LOCATION: 4059..4103

OTHER INFORMATION: /note= "exon 10"

FEATURE:

NAME/KEY: exon

LOCATION: 5543..5687

OTHER INFORMATION: /note= "exon 11"

FEATURE:

NAME/KEY: exon

LOCATION: 5688..5929

OTHER INFORMATION: /note= "putative.alt.exon"

; NAME/KEY: exon
; LOCATION: 7094...7221
; OTHER INFORMATION: /note= "exon 12"
US-08-870-126-11

Query Match 1.9%; Score 19; DB 3; Length 8310;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 697 GCCCGCGCGCGCTCGGC 715
|||||
Db 4280 GCCCGCGCGCGCTCGGC 4262

Search completed: April 12, 2003, 20:48:34
Job time : 139 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 19:17:08 ; Search time 2722 Seconds
(without alignments)
10467.174 Million cell updates/sec

Title: US-09-803-589-5
Perfect score: 979
Sequence: 1 gaattcgacagggccagc.....gaggatccctaccctcggc 979

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 15

Total number of hits satisfying chosen parameters: 28262

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vit.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	891	91.0	993	9	BC002851	BC002851 Homo sapi
2	840	85.8	998	6	AX083428	AX083428 Sequence
3	785	80.2	969	9	AF181562	AF181562 Homo sapi
4	678	69.3	780	6	AX083418	AX083418 Sequence
5	375	38.3	113853	9	AF196971	AF196971 Homo sapi
6	286	23.1	301	6	AX336215	AX336215 Sequence
7	226	23.1	301	6	AX409054	AX409054 Sequence
8	32	3.3	945	10	AF181561	AF181561 Rattus no
9	32	3.3	975	10	BC012263	BC012263 Mus muscu
10	32	3.3	1025	10	AF293356	AF293356 Mus muscu
11	32	3.3	1027	10	AF181560	AF181560 Mus muscu
12	32	3.3	62396	2	AC098339	AC098339 Rattus no
13	32	3.3	77876	2	AC098473	AC098473 Rattus no
14	32	3.3	111948	2	AC094563	AC094563 Rattus no
15	32	3.3	151694	10	AL670169	AL670169 Mouse DNA
16	32	3.3	177613	2	AC123014	AC123014 Rattus no
17	32	3.3	177613	2	AC123014	AC123014 Rattus no
18	25	2.6	1427	5	CHRW4P	D31900 Chicken mRN
19	24	2.5	109521	2	AP005551	AP005551 Oryza sat
20	22	2.2	425	6	AX071185	AX071185 Sequence
21	22	2.2	501	10	AF155166	AF155166 Rattus no
22	22	2.2	1519	10	AF064874	AF064874 Mus muscu
23	22	2.2	3655	10	AF332063	AF332063 Mus muscu
24	22	2.2	3656	10	AF332064	AF332064 Mus muscu
25	22	2.2	3993	10	AF247453	AF247453 Rattus no
26	22	2.2	167446	2	OSJN00085	AL606682 Oryza sat
27	22	2.2	172152	2	AC115188	AC115188 Rattus no
28	22	2.2	176097	2	AC112282	AC112282 Rattus no
29	22	2.2	204050	1	AL646070	AL646070 Ralstonia
30	22	2.2	214331	2	AL732312	AL732312 Mus muscu
31	22	2.2	220181	2	AC094816	AC094816 Rattus no
32	22	2.2	239758	10	AL591495	AL591495 Mouse DNA
33	22	2.2	257596	2	AC093408	AC093408 Mus muscu
34	21	2.1	533	6	AX105591	AX105591 Sequence
35	21	2.1	3737	4	AB022927	AB022927 Streptococ
36	21	2.1	7365	1	STMSVIDE0X	L37334 Streptomyce
37	21	2.1	39250	1	SVI011500	AP003864 Oryza sat
38	21	2.1	51661	8	AP003864	AP003864 Oryza sat
39	21	2.1	97310	2	AC103083	AC103083 Rattus no
40	21	2.1	139814	2	AP003964	AP003964 Oryza sat
41	21	2.1	154918	9	AL138781	AL138781 Human DNA
42	21	2.1	162297	2	AC108312	AC108312 Rattus no
43	21	2.1	179931	2	AC123246	AC123246 Rattus no
44	21	2.1	202429	2	AC130160	AC130160 Rattus no
45	20	2.0	243	4	BIN421636	AJ421636 Bos indic

ALIGNMENTS

RESULT 1
BC002851

LOCUS

DEFINITION

ACCESION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

BC002851 Homo sapiens, granin-like neuroendocrine peptide precursor, clone
MGC:3466 IMAGE:3660001, mRNA, complete cds.

BC002851

BC002851.1 GI:12804000

MGC.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 993)

Strausberg,R.

Direct Submission

993 bp mRNA linear PRI 12-JUL-2001

JOURNAL

Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Galithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nih.gov

Benjamin, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,

Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAL Plate: 12 Row: n Column: 24
This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.

FEATURES
source

Location/Qualifiers

1..993

/organism="Homo sapiens"

/db_xref="LocusID:27344"

/db_xref="taxon:9606"

/clone="MGC:3466 IMAGE:3660001"

/tissue.type="Uterus; endometrium adenocarcinoma"

/clone.lib="NH_MGC_44"

/lab_host="DH10B-R"

/note="vector: pOTB7"

30..812

/codon_start=1

/product="granin-like neuroendocrine peptide precursor"

/protein_id="AAH02851.1"

/db_xref="GI:12804001"

/translation="MAGSPLLMGPRAGVGLVLLVLLGLFRPPALCARPVKEPRGLS

AASPLAETGAPRRFRSVPGEAAGVLAALAHLEAERQERARAEAEADQ

ARVLQRLVMGAPRNSDPALGLDDPDPAQAOLARLLRDLPAALQOLVPAPVP

AAALRPVPVDDGAPGAEAGDETPDVPDLLRLVLLGRILAGSADSEGVAPRRL

RRADHDVGSLELPEGLGALLRKRLLETPAPQVPARRLLPP"

136 a 381 c 343 g 133 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 91.0%; Score 891; DB 9; Length 993;

Matches 941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	38	GGCCGGCTCGTGGGCGAGATGGCGGGTCCCGCTCTCTGGGGGCGCGGGCGGGG	97
Db	10	GGCCGGCTCGTGGGCGAGATGGCGGGTCCCGCTCTCTGGGGGCGCGGGCGGGG	69
Qy	98	GGCTGGCGCTTTGGTGTCTGTCTGGCTCGGCTTTTCGGCGCGCGCGCGCTCTCG	157
Db	70	GGCTGGCGCTTTGGTGTCTGTCTGGCTCGGCTTTTCGGCGCGCGCGCGCTCTCG	129
Qy	158	CGCGCGCGGTAAGAGCGCGCGCTTAAGCGAGCTCTTCGGCGCTTGGCTGAGACTG	217
Db	130	CGCGCGCGGTAAGAGCGCGCGCTTAAGCGAGCTCTTCGGCGCTTGGCTGAGACTG	189
Qy	218	GGCTTCCTCGCGCTTCCGGCGTCTAGTGGCCCCCGAGGTGAGCGCGCGGGGTGACG	277
Db	190	GGCTTCCTCGCGCTTCCGGCGTCTAGTGGCCCCCGAGGTGAGCGCGCGGGGTGACG	249
Qy	278	AGCTGGCGGGGCGCTGCGCATCTGCTGGAGCGGACAGCTCAGGAGCGGGCGGGCGG	337

Db	250	AGCTGGCGGGGCGCTGCGCATCTGCTGGAGGCCCAAGCTCAGAGCGGGGCGGGCGG	309
Qy	338	AGGCCAGGAGGCTGAGGATCAGAGCGCGCTCTGGCGAGCTGCTGCCGCTGTGGG	397
Db	310	AGGCCAGGAGGCTGAGGATCAGAGCGCGCTCTGGCGAGCTGCTGCCGCTGTGGG	369
Qy	398	GGCCCCCGCAACTCTGATCCGGCTTGGGGCTTGGAGCAGACACCCCGCCCTGCAG	457
Db	370	GGCCCCCGCAACTCTGATCCGGCTTGGGGCTTGGAGCAGACACCCCGCCCTGCAG	429
Qy	458	GGCAGCTGCTGCGCTCTGCTCCGGCGCGCTTGGAGCAGACACCCCGCCCTGCAG	517
Db	430	GGCAGCTGCTGCGCTCTGCTCCGGCGCGCTTGGAGCAGACACCCCGCCCTGCAG	489
Qy	518	TTGTCCTCCCGCGCTTCCCGCGCGCTTCCGAGCCCGCGCGCTTACGAGCAGG	577
Db	490	TTGTCCTCCCGCGCTTCCCGCGCGCTTCCGAGCCCGCGCGCTTACGAGCAGG	549
Qy	578	GGCCCCCGCGCGCTGATGAGGAGGCGGAGCAGACACCCCGCCCTGCAGCAGC	637
Db	550	GGCCCCCGCGCGCTGATGAGGAGGCGGAGCAGACACCCCGCCCTGCAGCAGC	609
Qy	638	TGTTGAGTACTTGTGGAGCGGATCTTGGGGAAGCGGACTCCGAGGGGGTGGCAG	697
Db	610	TGTTGAGTACTTGTGGAGCGGATCTTGGGGAAGCGGACTCCGAGGGGGTGGCAG	669
Qy	698	CCCCCGCGCGCTTCCCGCGCGCTTCCGAGCCCGCGCGCTTCCGAGCCCGCGCT	757
Db	670	CCCCCGCGCGCTTCCCGCGCGCTTCCGAGCCCGCGCGCTTCCGAGCCCGCGCT	729
Qy	758	CGGTCTGGGGGCGCTGCTGTGAACGCTAGAGACCCCGCGCGCTTCCGAGCCCTG	817
Db	730	CGGTCTGGGGGCGCTGCTGTGAACGCTAGAGACCCCGCGCGCTTCCGAGCCCTG	789
Qy	818	CACGCGCGCTTGTGGCAGCTGCGCGGATCCCGGATCCCGGATCCCGGATCCCGG	877
Db	790	CACGCGCGCTTGTGGCAGCTGCGCGGATCCCGGATCCCGGATCCCGGATCCCGG	849
Qy	878	TGCCCCCGCATCCCGCGCGCTTCCCGCGCGCTTCCCGCGCGCTTCCGAGCACTTAC	937
Db	850	TGCCCCCGCATCCCGCGCGCTTCCCGCGCGCTTCCCGCGCGCTTCCGAGCACTTAC	909
Qy	938	GGCGCAGCAGCCCTCTCACCCGAGGATCCCTTACCCCTTGGC	979
Db	910	GGCGCAGCAGCCCTCTCACCCGAGGATCCCTTACCCCTTGGC	951

RESULT 2
AX083428
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

AX083428
Sequence 120 from Patent WO0112660.
AX083428
AX083428.1 GI:13185274
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 998)
Kato, S. and Kimura, T.
Human proteins having hydrophobic domains and dnas encoding these
proteins
Patent: WO 0112660-A 120 22-FEB-2001;
SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
Location/Qualifiers
1..998
/organism="Homo sapiens"
/db_xref="taxon:9606"
50..832
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC33311.1"

```
/db_xref="GI:13185275"
/translation="MAGSPLLMGPRAGGVGLLVLLGLFRPPPALCARVKEPRGLS
AASPLAETGAPRRFRSYPGEAGVQELARALAHLEAEQERARAEAEQ
ARVLAQLLRVWGPARNSDPALGDDPDPAQAQALRALRLDPAALAAQVLPVP
AAALRPVPYDDGPAQDAEAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRL
RAADHDVGSSELPPEGLGALLRVKRLTTPAPQVPARLLPP"
BASE COUNT      120 a 394 c 347 g 137 t
ORIGIN
Query Match      85.8%; Score 840; DB 6; Length 998;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 940; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 38 GCCCGGCTCGCTGGGCGAGCATGGCGGGTCGCCGCTGCTCTGGGGGCCCGGGCGGGG 97
Db 30 GCCCGGCTCGCTGGGCGAGCATGGCGGGTCGCCGCTGCTCTGGGGGCCCGGGCGGGG 89
QY 98 GCGTCGCGCTTTTGGTGTCTGCTGCTGGCGCTGTTTGGCGCGCCCGCGCGCTCTGCG 157
Db 90 GCGTCGCGCTTTTGGTGTCTGCTGCTGGCGCTGTTTGGCGCGCCCGCGCGCTCTGCG 149
QY 158 CGCGCGCGGTAAAGGAGCCCGCGGCGCTAAGCGCAGCGTCTCGCGCCCTTGGCTGAGACTG 217
Db 150 CGCGCGCGGTAAAGGAGCCCGCGGCGCTAAGCGCAGCGTCTCGCGCCCTTGGCTGAGACTG 209
QY 218 GCGCTCTCTCGCGCTTCCGCGGCTCAGTCCCGGAGGTGAGCGCGGGGGCGGCTGCAGG 277
Db 210 GCGCTCTCTCGCGCTTCCGCGGCTCAGTCCCGGAGGTGAGCGCGGGGGCGGCTGCAGG 269
QY 278 AGCTGCGCGCGGCGTGGCGCATCTGCTGAGAGCCGACGCTCAGGAGCGGGCGGGCGCG 337
Db 270 AGCTGCGCGCGGCGTGGCGCATCTGCTGAGAGCCGACGCTCAGGAGCGGGCGGGCGCG 329
QY 338 AGCGCAGGAGGCTGAGGATCAGCAGCGCGCGCTCTGCGCGAGCTGCTGCGCGTCTGGG 397
Db 330 AGCGCAGGAGGCTGAGGATCAGCAGCGCGCGCTCTGCGCGAGCTGCTGCGCGTCTGGG 389
QY 398 GCGCGCCCGCCCACTCTGATCCGGCTCTGGGCTTGAGCAGCAGCACCOCGAGCGCCCTGCAG 457
Db 390 GCGCGCCCGCCCACTCTGATCCGGCTCTGGGCTTGAGCAGCAGCACCOCGAGCGCCCTGCAG 449
QY 458 CGCAGCTCGCTCGCTGCTCTCGCGCGCGCTTACCCCTGCGCGCCCTAGCAGCCCGCAGC 517
Db 450 CGCAGCTCGCTCGCTGCTCTCGCGCGCGCTTACCCCTGCGCGCCCTAGCAGCCCGCAGC 509
QY 518 TTGTCCCGCGCGCGTCCCGCGCGCGCTCCGAGACCCCGCGCGCGTCTAGCAGGACG 577
Db 510 TTGTCCCGCGCGCGTCCCGCGCGCGCTCCGAGACCCCGCGCGCGTCTAGCAGGACG 569
QY 578 GCGCGCGCGCGCGGATGCTGAGGAGCGAGCGGCGAGCAGACACCAGCTGGAGCCCGCAGC 637
Db 570 GCGCGCGCGCGCGGATGCTGAGGAGCGAGCGGCGAGCAGACACCAGCTGGAGCCCGCAGC 629
QY 638 TGTGTAGGTACTTGTGGGAGCGATCTTTCGGGAGAGCGGCGACTCCGAGGGGGTGGCAG 697
Db 630 TGTGTAGGTACTTGTGGGAGCGATCTTTCGGGAGAGCGGCGACTCCGAGGGGGTGGCAG 689
QY 698 CCGCGCGCGCGCTCCGCGCTGCGCGCGACACCATGTGGGCTCTGAGCTGCCCGCTGAGG 757
Db 690 CCGCGCGCGCGCTCCGCGCTGCGCGCGACACCATGTGGGCTCTGAGCTGCCCGCTGAGG 749
QY 758 GCGTGTGCGGGGCGTGTCTGCTGTGAACCGCTTAGAGACCCCGCGCGCGCCAGGTGCCTG 817
Db 750 GCGTGTGCGGGGCGTGTCTGCTGTGAACCGCTTAGAGACCCCGCGCGCGCCAGGTGCCTG 809
QY 818 CAGCGCGCGCTTGTGCACCTGAGCACTGCCCGGATCCCGTGCACCTTGGGAGCCCGAGAAG 877
Db 810 CAGCGCGCGCTTGTGCACCTGAGCACTGCCCGGATCCCGTGCACCTTGGGAGCCCGAGAAG 869
QY 878 TGCCCCCGGCGCATCCCGCGCAGGACTGTCTCCCGCGCAGCAGCTCCAGAGCAACTTACCC 937
Db 870 TGCCCCCGGCGCATCCCGCGCAGGACTGTCTCCCGCGCAGCAGCTCCAGAGCAACTTACCC 929
```

```
QY 938 CGGCCAGCCAGCCCTCTACCCGAGGATCCCTACCCCTCGGC 979
Db 930 CGGCCAGCCAGCCCTCTACCCGAGGATCCCTACCCCTCGGC 971
RESULT 3
AF181562 969 bp mRNA linear PRI 27-JAN-2000
LOCUS Homo sapiens proSAAS mRNA, complete cds.
DEFINITION AF181562
ACCESSION AF181562
VERSION AF181562.1 GI:6653212
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 969)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devil,L.A. and Douglass,J.
IDENTIFICATION AND CHARACTERIZATION OF PROSAAS, A GRANIN-LIKE
NEUROENDOCRINE PEPTIDE PRECURSOR THAT INHIBITS PROHORMONE
PROCESSING
J. Neurosci. 20 (2), 639-648 (2000)
JOURNAL 20098938
MEDLINE 10632593
PUBMED 10632593
REFERENCE 2 (bases 1 to 969)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devil,L.A. and Douglass,J.
DIRECT SUBMISSION
Submitted (30-AUG-1999) Department of Molecular Pharmacology,
Albert Einstein College of Medicine, 1300 Morris Park Avenue,
Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="FLHSAAS"
1..26
5'UTR
27..809
CDS
/function="inhibits prohormone processing"
/note="granin-like neuroendocrine peptide precursor"
/codon_start=1
/product="proSAAS"
/protein_id="AAF22643.1"
/db_xref="GI:6653213"
/translation="MAGSPLLMGPRAGGVGLLVLLGLFRPPPALCARVKEPRGLS
AASPLAETGAPRRFRSYPGEAGVQELARALAHLEAEQERARAEAEQ
ARVLAQLLRVWGPARNSDPALGDDPDPAQAQALRALRLDPAALAAQVLPVP
AAALRPVPYDDGPAQDAEAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRL
RAADHDVGSSELPPEGLGALLRVKRLTTPAPQVPARLLPP"
sig_peptide 27..125
3'UTR 810..969
BASE COUNT 117 a 379 c 338 g 135 t
ORIGIN
Query Match 80.2%; Score 785; DB 9; Length 969;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 935; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 42 GGCTCGTGGGCGAGCATGGCGGGTGGCGCTGCTCTGGGGCGCGCGGGCGGCGGT 101
Db 11 GGCTCGTGGGCGAGCATGGCGGGTGGCGCTGCTCTGGGGCGCGCGGGCGGCGGT 70
QY 102 CGGCTTTTGGTGTGCTGCTGCTCGGCTGTTTGGCGCGCGCGCGCGCTCTGCGCGCG 161
Db 71 CGGCTTTTGGTGTGCTGCTGCTCGGCTGTTTGGCGCGCGCGCGCGCTCTGCGCGCG 130
QY 162 GCGCGTAAAGGAGCCCGCGGCTTAGCGCAGCGTCTCCGCCCTTGGCTGAGACTGGCGC 221
Db 131 GCGCGTAAAGGAGCCCGCGGCTTAGCGCAGCGTCTCCGCCCTTGGCTGAGACTGGCGC 190
```

Qy	222	TCCTCCGCGCTTCCGCGGTTCAGTCCCGCAGGTCAGCGGGGGGGGGGTCAGGAGCT	281
Db	191	TCCTCCGCGCTTCCGCGGTTCAGTCCCGCAGGTCAGCGGGGGGGGGTTCAGGAGCT	250
Qy	282	GGCGGGGCGCTGGCGCATCTGCTGGAGCGCAAGCTCAGAGCGGGCGCGGGCGAGGC	341
Db	251	GGCGGGGCGCTGGCGCATCTGCTGGAGCGCAAGCTCAGAGCGGGCGCGGGCGAGGC	310
Qy	342	CAGAGAGCTGAGGATCAGCAGCGCGCTCCTTGCGCAGCTGCTGCGCTTGGGGCGC	401
Db	311	CAGAGAGCTGAGGATCAGCAGCGCGCTCCTTGCGCAGCTGCTGCGCTTGGGGCGC	370
Qy	402	CCCCCGCAACTCTGATCCGGCTCTGGGCTTGAGACAGACACCCGACGCGCTGCAGCGCA	461
Db	371	CCCCCGCAACTCTGATCCGGCTCTGGGCTTGAGACAGACACCCGACGCGCTGCAGCGCA	430
Qy	462	GCTCGCTCGCGCTCTGCTCCGCGCGCTTGACCTCCGCGCTTACGAGCCAGCTTGT	521
Db	431	GCTCGCTCGCGCTCTGCTCCGCGCGCTTGACCTCCGCGCTTACGAGCCAGCTTGT	490
Qy	522	CCCCCGCGCTTCCCGCGCGGCGCTCCGACCCCGCGCGCTTACGAGCGAGCGGCC	581
Db	491	CCCCCGCGCTTCCCGCGCGGCGCTCCGACCCCGCGCGCTTACGAGCGAGCGGCC	550
Qy	582	CGCGGCGCGGATGCTGAGGAGCGCAGCGACAGACACCCGACGCTGGACCGAGCTGTT	641
Db	551	CGCGGCGCGGATGCTGAGGAGCGCAGCGACAGACACCCGACGCTGGACCGAGCTGTT	610
Qy	642	GAGGTACTTGTGGACGATTTCTGCGGGAAGCGCGACCTCCGAGGGGTGGCAGCGCC	701
Db	611	GAGGTACTTGTGGACGATTTCTGCGGGAAGCGCGACCTCCGAGGGGTGGCAGCGCC	670
Qy	702	CGCGCGCTTCCGCGCTGCGCGCACACGATGTGGGCTCTGAGCTGCCCCCTGAGGGGT	761
Db	671	CGCGCGCTTCCGCGCTGCGCGCACACGATGTGGGCTCTGAGCTGCCCCCTGAGGGGT	730
Qy	762	GCTGGGGCGCTGCTGCTGTGAACGCTAGAGACCCCGCGCGCTGAGCTGCGCTGCACG	821
Db	731	GCTGGGGCGCTGCTGCTGTGAACGCTAGAGACCCCGCGCGCTGAGCTGCGCTGCACG	790
Qy	822	CCGCTCTTTCACACCTGAGCACTGCCCGGATCCGTCACCTGGGACCCAGAGTGCC	881
Db	791	CCGCTCTTTCACACCTGAGCACTGCCCGGATCCGTCACCTGGGACCCAGAGTGCC	850
Qy	882	CCGCGCATCCCGCACAGGACTGCTCCCGCGCAGCAGCTCCAGAGCAACTTACCCCGGC	941
Db	851	CCGCGCATCCCGCACAGGACTTCTCCCGCGCAGCAGCTCCAGAGCAACTTACCCCGGC	910
Qy	942	CAGCCAGCCCTCTCACCCGAGGATCCCTTACCCCTGGC	979
Db	911	CAGCCAGCCCTCTCACCCGAGGATCCCTTACCCCTGGC	948

RESULT 4	AX083418	780 bp	DNA	linear	PAT 28-FEB-2001
LOCUS	Sequence 110 from Patent WO0112660.				
DEFINITION	AX083418				
ACCESSION	AX083418.1	GI:13185255			
VERSION					
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 780)				
AUTHORS	Kato, S. and Kimura, T.				
TITLE	Human proteins having hydrophobic domains and dnas encoding these proteins				
JOURNAL	Patent: WO 0112660-A 110 22-FEB-2001;				
FEATURES	SAGAMI CHEMICAL RESEARCH CENTER (JP); Protegene Inc. (JP)				
Location/Qualifiers	1..780				
source	/organism="Homo sapiens"				

BASE COUNT	79 a	297 c	294 g	110 t	
ORIGIN					
Query Match	69.3%	Score 678:	DB 6:	Length 780;	
Best Local Similarity	99.7%	Pred. No. 0;			
Matches 778:	Conservative	0;	Mismatches	2: Indels 0: Gaps 0:	
QY	58	ATGCGGGGTCGCGCTGCTCTGGGGGCGCGGGGCGGGGGCGTGGGCTCTTGGTGTGCTG	117		
Db	1	ATGCGGGGTCGCGCTGCTCTGGGGGCGCGGGGCGGGGGCGTGGGCTCTTGGTGTGCTG	60		
QY	118	CTGTGCTGCGCGCTGTTTCGGGCGCGCGCGCGCTCTGCGCGCGCGGTAAGAGAGCC	177		
Db	61	CTGTGCTGCGCGCTGTTTCGGGCGCGCGCGCGCTCTGCGCGCGCGGTAAGAGAGCC	120		
QY	178	CGGGGCTAAGCGCAGCGTCTCGGCCCTTGGCTGAGACTGGGCTCTCTCGCGCTCCGG	237		
Db	121	CGGGGCTAAGCGCAGCGTCTCGGCCCTTGGCTGAGACTGGGCTCTCTCGCGCTCCGG	180		
QY	238	CGGTACGTGCGCGAGTGAGGCGGGGCGGTGCGAGAGCTGGCGGGCGGCTGGCG	297		
Db	181	CGGTACGTGCGCGAGTGAGGCGGGGCGGTGCGAGAGCTGGCGGGCGGCTGGCG	240		
QY	298	CATCTGCTGAGGCGCAAGCTCAGGAGCGGGCGGGCGAGGCGCAGAGGCTGAGGAT	357		
Db	241	CATCTGCTGAGGCGCAAGCTCAGGAGCGGGCGGGCGCAGGCGCAGAGGCTGAGGAT	300		
QY	358	CAGCAGCGCGGCTCTGCGGCGAGCTGCTGGCGGCTTGGGGGCGCCCCCGCAACTCTGAT	417		
Db	301	CAGCAGCGCGGCTCTGCGGCGAGCTGCTGGCGGCTTGGGGGCGCCCCCGCAACTCTGAT	360		
QY	418	CCGGCTCTGGGCTTGACAGCAGACCCGACGCGCTGACGCGAGCTGCTCGCGCTCTG	477		
Db	361	CCGGCTCTGGGCTTGACAGCAGACCCGACGCGCTGACGCGAGCTGCTCGCGCTCTG	420		
QY	478	CTCCGGCGCGGCTTGACCTTGCCTGCGGCTTACAGCGCGAGCTTGTCCCGCGCGCTCCG	537		
Db	421	CTCCGGCGCGGCTTGACCTTGCCTGCGGCTTACAGCGCGAGCTTGTCCCGCGCGCTCCG	480		
QY	538	GCGCGGGGCTCGGACCCGCGGCTTACGAGCAGCGCGCGCGCGCGCGGCGGATGCT	597		
Db	481	GCGCGGGGCTCGGACCCGCGGCTTACGAGCAGCGCGCGCGCGCGCGGCGGATGCT	540		
QY	598	GAGGAGCGAGGCGACGAGACACCGAGCTGGACCCCGAGCTGTGAGGTACTTGTGTGGGA	657		
Db	541	GAGGAGCGAGGCGACGAGACACCGAGCTGGACCCCGAGCTGTGAGGTACTTGTGTGGGA	600		
QY	658	CGGATTTCTTGGGGAAGCGGACTCCGAGGGGTGGCAGCGCGCGCGCTTGGCGCGT	717		
Db	601	CGGATTTCTTGGGGAAGCGGACTCCGAGGGGTGGCAGCGCGCGCGCTTGGCGCGT	660		
QY	718	GCGCGCGCAGCATGTGGGCTCTGAGTGCCTTGGGCGGCTGTTGGGGCGGCTGCTG	777		
Db	661	GCGCGCGCAGCATGTGGGCTCTGAGTGCCTTGGGCGGCTGTTGGGGCGGCTGCTG	720		
QY	778	CGTGTGAACGCTTAGAGACCCGCGCGCGCTGACGCGCGCTTGGCGCGCTTGGCACCC	837		
Db	721	CGTGTGAACGCTTAGAGACCCGCGCGCGCTGACGCGCGCTTGGCGCGCTTGGCACCC	780		
RESULT 5					
LOCUS	AF196971/c	113853 bp	DNA	linear	PRI 09-NOV-1999
DEFINITION	Homo sapiens GATA-binding protein 1 and histone deacetylase-like protein genes, complete cds; CRAS pseudogene, complete sequence; and protein translocase gene, partial cds.				
ACCESSION	AF196971				
VERSION	AF196971.1 GI:6289080				
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

```

REFERENCE      Mammalia: Eutheria: Primates; Catarrhini; Hominiidae; Homo.
AUTHORS        1 (bases 1 to 113853)
               Bleischmidt, K., Nyakatura, G., Strom, T.M., Drescher, B., Menzel, U.,
               Meindl, A. and Rosenthal, A.
TITLE          Direct Submission
JOURNAL        Submitted (14-OCT-1999) Genome Analysis, Institute of Molecular
               Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT        1-36552: cosmid J1045; 11106-52106: cosmid E2237; 27017-67034:
               cosmid L1850; 58536-92644: cosmid D1425; 80573-113853: cosmid
               E1239.
FEATURES       Location/Qualifiers
               1. .113853
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosomes="X"
               /map="Xp11.23"
               /clone="cosmids J1045; E2237; L1850; D1425; E1239"
               1. .252
               /evidence=not_experimental
               /rpt_family="AluSx"
               254. .396
               /evidence=not_experimental
               /rpt_family="L1M4"
               397. .704
               /evidence=not_experimental
               /rpt_family="AluSg"
               705. .717
               /evidence=not_experimental
               /rpt_family="L1M4"
               817. .1083
               /evidence=not_experimental
               /rpt_family="L1MB7"
               complement(1084. .1371)
               /evidence=not_experimental
               /rpt_family="AluJo"
               1372. .1790
               /evidence=not_experimental
               /rpt_family="L1MB7"
               1802. .1924
               /evidence=not_experimental
               /rpt_family="L1MD2"
               complement(2269. .2440)
               /evidence=not_experimental
               /rpt_family="L2"
               complement(2950. .3059)
               /evidence=not_experimental
               /rpt_family="MIR"
               3105. .3235
               /evidence=not_experimental
               /rpt_family="MIR"
               3268. .3472
               /evidence=not_experimental
               /rpt_family="AluJo"
               3504. .3608
               /note="homology = 80.00%, score = 28, counts = 3"
               /evidence=not_experimental
               /rpt_type=tandem
               /rpt_unit=ttagcattagaatagtttactattataaattagta
               3613. .3955
               /evidence=not_experimental
               /rpt_family="L1MA5"
               3964. .4059
               /note="homology = 77.10%, score = 20, counts = 3"
               /evidence=not_experimental
               /rpt_type=tandem
               /rpt_unit=aataattagttactactaataagttacgtat
               complement(4121. .4369)
               /evidence=not_experimental
               /rpt_family="MIR"
               4393. .4463
               /evidence=not_experimental
               /rpt_family="L2"
               complement(4502. .4544)
               /evidence=not_experimental
               /rpt_family="MIR"
               /evidence=not_experimental
               /rpt_family="MIR"
               5062. .5157
               /note="homology = 67.70%, score = 32, counts = 48"
               /rpt_type=tandem
               /rpt_unit=tc
               5649. .5691
               /evidence=not_experimental
               /rpt_family="MIR"
               6131. .13747
               /gene="GATA-1"
               join(6131. .6168,10614. .10852,11367. .11744,11846. .11991,
               12695. .12820,13316. .13747)
               /gene="GATA-1"
               /product="GATA-binding protein 1"
               /note="mRNA acc. no. M30601"
               6131. .6168
               /gene="GATA-1"
               /number=1
               8548. .8857
               /evidence=not_experimental
               /rpt_family="AluSx"
               9694. .9979
               /evidence=not_experimental
               /rpt_family="AluJb"
               10614. .10852
               /gene="GATA-1"
               /number=2
               join(10633. .10852,11367. .11744,11846. .11991,12695. .12820,
               13316. .13687)
               /gene="GATA-1"
               /note="globin transcription factor 1; GF-1"
               /codon_start=1
               /product="GATA-binding protein 1"
               /protein_id="NAF06806.1"
               /db_xref="GI:6289081"
               /translation="MEFFGLSGTSELPQVDPALVSSPTESGVFFPSGPEGLDAA
               ASSTAPATAAAALAYRDAEAYRHSPFQVYPLNMGEGIPGSGPYAGWAYGKVTG
               LYPASTVCTREDSPQAVEDLDGKSTFLETKLTERLSPDLLTGLPALPSLPVPN
               SAYGGPDSFTSPGSPNSAAVSSPKLRTLPDPCAEARECVNCGATAPLWRRD
               RTGHVLCNAGLYHKMNGONRPLIRPKELIYSKAGTQCTNCOTTTTLWRNNSGD
               PNCAGGLIYKLHQVNRPLTMRKDGIOIRNRKASGKRGKSGSGGTAAGGAGGPF
               MVVAGSGNCGEVSAGSLTLPQGTATHLYQGLGVPVLVSGPVSHLMPFPGPLLGSPTG
               SPTGPMPTTSTTVVPLSS"
               11367. .11744
               /gene="GATA-1"
               /number=3
               11846. .11991
               /gene="GATA-1"
               /number=4
               complement(12330. .12477)
               /evidence=not_experimental
               /rpt_family="MER5B"
               12695. .12820
               /gene="GATA-1"
               /number=5
               13316. .13747
               /gene="GATA-1"
               /number=6
               14173. .14226
               /note="homology = 98.10%, score = 25, counts = 2"
               /evidence=not_experimental
               /rpt_type=tandem
               /rpt_unit=ggcatgacgtggggacaccttttagac
               14290. .14375
               /note="homology = 62.80%, score = 20, counts = 43"
               /evidence=not_experimental
               /rpt_type=tandem
               /rpt_unit=gt
               complement(14876. .14972)
               /evidence=not_experimental
               /rpt_family="MIR"

```

```
repeat_region complement(14899..15178)
/evidence-not_experimental
/rpt_family="MIR"
16160..16259
/note="homology = 74.00%, score = 23, counts = 4"
/evidence-not_experimental
/rpt_type=tandem
/rpt_unit=atacaggaacccctgtctcaccat
16338..16674
/evidence-not_experimental
/rpt_family="MER1B"
16780..16836
/note="homology = 86.00%, score = 22, counts = 3"
/evidence-not_experimental
/rpt_type=tandem
/rpt_unit=ttatttacaggaatctctg
16853..17048
/note="homology = 75.00%, score = 24, counts = 4"
/evidence-not_experimental
/rpt_type=tandem
/rpt_unit=ctactgtctgaacccacaaggaa
complement(17629..17940)
/evidence-not_experimental
/rpt_family="MER33"
18028..18353
/evidence-not_experimental
/rpt_family="MLT1I"
complement(18388..18668)
/evidence-not_experimental
/rpt_family="AluSx"
complement(18674..18977)
/evidence-not_experimental
/rpt_family="AluSg"
19143..19448
/evidence-not_experimental
/rpt_family="AluDb"
19750..19923
/evidence-not_experimental
/rpt_family="MER58C"
complement(20505..20587)
/evidence-not_experimental
/rpt_family="MIR"
20598..20658
/evidence-not_experimental

Query Match 38.3%; Score 375; DB 9; Length 113853;
Best Local Similarity 99.6%; Pred. No. 2.2e-170;
Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 170 AGGAGCCCGCGGCTAAGCGACGCTCGCCCTTGGCTGAGACTGGCGCTCCTCGCC 229
DB 51869 AGGAGCCCGCGGCTAAGCGACGCTCGCCCTTGGCTGAGACTGGCGCTCCTCGCC 51810

QY 230 GCTTCGGCGGTGAGTCCCGGAGGTGAGCGCGCGGGGGGGGTGAGAGTGGCGGG 289
DB 51809 GCTTCGGCGGTGAGTCCCGGAGGTGAGCGCGCGGGGGGGGTGAGAGTGGCGGG 51750

QY 290 CGCTGGCGCATCTGCTGGAGGCCGAAGCTCAGGAGCGGGCGCGCGCGGCGAGGAGG 349
DB 51749 CGCTGGCGCATCTGCTGGAGGCCGAAGCTCAGGAGCGGGCGCGCGGCGAGGAGG 51690

QY 350 CTGAGGATCAGACGGCGGCTCTGCGGAGCTGCTGCGGCTTGGGGGCGCCCGCCGA 409
DB 51689 CTGAGGATCAGACGGCGGCTCTGCGGAGCTGCTGCGGCTTGGGGGCGCCCGCA 51630

QY 410 ACTCTGATCCGGCTTGGCTTGACACGACGACCCGCGGCTGACGCGAGTCCGCTC 469
DB 51629 ACTCTGATCCGGCTTGGCTTGACACGACGACCCGCGGCTGACGCGAGTCCGCTC 51570

QY 470 GCGCTCTGCTCCGCGCGGCTTACCCCTGCGGCTAGACGCGGCTGCTGCTCCCGCGC 529
DB 51569 GCGCTCTGCTCCGCGCGGCTTACCCCTGCGGCTGCGGCGGCTGCTGCTCCCGCGC 51510
```

```
QY 530 CCGTCCCGCGCGCGCTCCGACCCCGCGCGCGCGCGTCTACGACGACGCGCGCGGCGCC 589
DB 51509 CCGTCCCGCGCGCGCGCTCCGACCCCGCGCGCGCGCGTCTACGACGACGCGCGCGGCGCC 51450

QY 590 CGGATGCTGAGGAGCGGCGGACGACGACGCGGAGTGGACCGCGGAGTGTGAGGT 646
DB 51449 CGGATGCTGAGGAGCGGCGGACGACGACGCGGAGTGGACCGCGGAGTGTGAGGT 51393

RESULT 6
AX336215/c 301 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 6724 from Patent WO0194629.
DEFINITION AX336215
ACCESSION AX336215.1 GI:18126934
VERSION AX336215.1
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 6724 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
source 1..301
/db_xref="taxon:9606"
BASE COUNT 41 a 79 c 131 g 48 t 2 others
ORIGIN

Query Match 23.1%; Score 226; DB 6; Length 301;
Best Local Similarity 99.6%; Pred. No. 3.6e-98;
Matches 276; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 703 CGCGCGCTCCGCGTGGCGGACGACGATGTGGGTCTGAGTGGCGCGCGGAGGCGGTG 762
DB 301 CGCGCGCTCCGCGTGGCGGACGACGATGTGGGTCTGAGTGGCGCGCGGAGGCGGTG 242

QY 763 CTGGGGCGGTGCTGCTGTGTAAGCGCTAGAGACCGCGCGCGCGCGGAGTGCCTGCAGCG 822
DB 241 CTGGGGCGGTGCTGCTGTGTAAGCGCTAGAGACCGCGCGCGCGGAGTGCCTGCAGCG 182

QY 823 CGCTCTTGCCACCGCTGAGCACTGCCGGATCCCGTGCACCGTGGGACCGAGAGTGGCC 882
DB 181 CGCTCTTGCCACCGCTGAGCACTGCCGGATCCCGTGCACCGTGGGACCGAGAGTGGCC 122

QY 883 CGGCATCCCGCCACCGAGGAGTGTCCCGGACGACGCTCCAGAGCACTTACCCCGGCGC 942
DB 121 CGGCATCCCGCCACCGAGGAGTGTCCCGGACGACGCTCCAGAGCACTTACCCCGGCGC 62

QY 943 AGCCAGCGCTCTACCCCGAGGATCCCTACCCCGTGGC 979
DB 61 AGCCAGCGCTCTACCCCGAGGATCCCTACCCCGTGGC 25

RESULT 7
AX409054/c 301 bp DNA linear PAT 14-JUN-2002
LOCUS Sequence 1701 from Patent WO0229103.
DEFINITION AX409054
ACCESSION AX409054
VERSION AX409054.1 GI:21441759
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE Gene expression profiles in liver cancer
```

JOURNAL Patent: WO 0229103-A 1701 11-APR-2002;

GENE LOGIC INC (US)

Location/Qualifiers

FEATURES

source

1..301

/organism="Homo sapiens"

/db_xref="taxon:9606"

/note="EMBL/GenBank Accession No. D59847"

BASE COUNT 41 a 79 c 131 g 48 t 2 others

ORIGIN

Query Match 23.1%; Score 226; DB 6; Length 301;

Best Local Similarity 99.6%; Pred. No. 3.6e-98;

Matches 276; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 703 CGCGCGCTCCGCGCGCCGACACGATGTGGGTCTGAGCTGCCCTGAGGCGGTG 762

Db 301 CGCGCGCTCCGCGCGCCGACACGATGTGGGTCTGAGCTGCCCTGAGGCGGTG 242

QY 763 CTGGGGCGCTGCTGGTGTGAACGCTAGAGACCCGCGCCGCTGCTGCACGC 822

Db 241 CTGGGGCGCTGCTGGTGTGAACGCTAGAGACCCGCGCCGCTGCTGCACGC 182

QY 823 CGCCTCTTCCACCTTGACACTGCCCGGATCCCTGACCTGGACCCAGAGTGCCC 882

Db 181 CGCCTCTTCCACCTTGACACTGCCCGGATCCCTGACCTGGACCCAGAGTGCCC 122

QY 883 CGGCATCCCGCCACAGACTGCTCCCGCCGACGACGCTCCAGACCACTTACCCGCGCC 942

Db 121 CGGCATCCCGCCACAGACTGCTCCCGCCGACGACGCTCCAGACCACTTACCCGCGCC 62

QY 943 AGCCAGCCTCTACCCGAGGATCCCTACCCCTGGC 979

Db 61 AGCCAGCCTCTACCCGAGGATCCCTACCCCTGGC 25

RESULT 8

AF181561

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

/note="granin-like neuroendocrine peptide precursor"

/codon_start=1

/product="proSAAS"

/protein_id="AAF22642.1"

/db_xref="GI:6653211"

/translation="MAGSPILCCGPRAGVGILLVLLGLLRLPTLSARPYKEPRSL

AASPLAETSPLKRLRAVPRGAVGAVQELARALAHLEERQERARAEAEEDQQ

ARVLAQLLRAGWSPRASDPLAPDDPAPAAQALRLARLDPAALAAQLVPAPAP

AAALRPVPVDDGPTGDEDADETDPDPELRLYLGLRILTGSSPEEAPAPRRL

RRAVDQDLGPEVPPENVLGLLRKLENSSPQAPARLLPP"

sig_peptide

3'UTR

polyA_signal

polyA_site

BASE COUNT

ORIGIN

134 a 355 c 307 g 149 t

Query Match 3.3%; Score 32; DB 10; Length 945;

Best Local Similarity 100.0%; Pred. No. 0.00023;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 TGGGGCAGCATGGCGGGTGGCGCTGCTCTG 80

Db 13 TGGGGCAGCATGGCGGGTGGCGCTGCTCTG 44

RESULT 9

BC012263

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Strausberg,R.

Direct Submission

Submitted (06-AUG-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: ang@bcm.tmc.edu

Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,

Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LILNL at: <http://image.llnl.gov>

Series: IRAK Plate: 24 Row: p Column: 24.

Location/Qualifiers

1..975

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="FVB/N"

/clone="MGC:19107 IMAGE:4207854"

/tissue_type="Colon, normal. 5 month old male mouse."

/clone_lib="NCI_CGAP_C024"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

FEATURES

source

1..945

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone="FLRSAAS"

1..21

22..804

/function="inhibits prohormone processing"

5'UTR

CDS

CDS 20..796
/codon_start=1
/product="Similar to granin-like neuroendocrine peptide precursor"
/protein_id="AAH12263.1"
/translation="MAGSPLLCGPRAGGVGILVLLGLLRLPPTLSARPVKEPRSL
AASAPLVETSTPLRLRAVRPGEAGAVQELARALAHLEAEQERARAQAEQDQ
ARVLAQLLRAGSPRASDPPLAPDDPDPAQAOLARALLRLDPAALAAQLVPAPAA
APRPVPVYDDGPTGPDVEDAGDETDVDPPELLRYLLGRITLTGSSEPEAPAPRLRR
SVDQDLGPEVPPENVLGLLRKLENPSQAPARLLPP"
BASE COUNT 173 a 353 c 304 g 145 t
ORIGIN
Query Match 3.3%; Score 32; DB 10; Length 975;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 49 TGGGGCAGCATGGCGGGTGCCTGCTCTG 80
|||||
Db 11 TGGGGCAGCATGGCGGGTGCCTGCTCTG 42
|||||
RESULT 10
AF293356 1025 bp mRNA linear ROD 31-AUG-2000
LOCUS Mus musculus IAA mRNA, complete cds.
DEFINITION AF293356
ACCESSION AF293356.1 GI:9954415
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1025)
AUTHORS Donadel,G., Marinos,N., DeSilva,M.G., Lu,J., Notkins,A.L. and Lan,M.S.
TITLE Molecular cloning and characterization of a highly basic protein, IA-4, expressed in pancreatic islets and brain
JOURNAL Neuroendocrinology 67 (3), 190-196 (1998)
MEDLINE 9630436
PUBMED
REFERENCE 2 (bases 1 to 1025)
AUTHORS Donadel,G., Notkins,A.L. and Lan,M.S.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2000) Laboratory of Oral Medicine, National Institute of Dental Research, 9000 Rockville Pike, Building 30 Room 124, Bethesda, MD 20892, USA
FEATURES
source
1..1025
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="insulinoma b-TC-1"
94..557
/notes="highly basic protein; highly expressed in pancreatic islets and neuronal cells"
/codon_start=1
/product="IA4"
/protein_id="AAG09048.1"
/db_xref="GI:9954416"
/translation="MAGSPLLCGPRAGGVGILVLLGLLRLPPTLSARPVKEPRSL
ARSAPLVETSTPLRLRAVRPGEAGAVQELARALAHLEAEQERARAQAEQDQ
RMSWRSSGAGALVRPRTPLAPDDPDPAQAOLARALLRLDPAALAAQLVPAPAA
PAPRPVPVYDDGPHWRRCRRRDW"
BASE COUNT 156 a 382 c 330 g 157 t
ORIGIN
Query Match 3.3%; Score 32; DB 10; Length 1025;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 49 TGGGGCAGCATGGCGGGTGCCTGCTCTG 80
|||||

Db 85 TGGGGCAGCATGGCGGGTGCCTGCTCTG 116
RESULT 11
AF181560 1027 bp mRNA linear ROD 27-JAN-2000
LOCUS Mus musculus proSAAS gene, complete cds.
DEFINITION AF181560
ACCESSION AF181560.1 GI:6653208
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1027)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L., Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N., Devi,L.A. and Douglass,J.
TITLE Identification and characterization of proSAAS, a granin-like neuroendocrine peptide precursor that inhibits prohormone processing
JOURNAL J. Neurosci. 20 (2), 639-648 (2000)
MEDLINE 20098938
PUBMED 10632593
REFERENCE 2 (bases 1 to 1027)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L., Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N., Devi,L.A. and Douglass,J.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1999) Department of Molecular Pharmacology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
FEATURES
source
1..1027
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="FLMSAAS"
1..109
110..886
/function="inhibits prohormone processing"
/note="granin-like neuroendocrine peptide precursor"
/codon_start=1
/product="proSAAS"
/protein_id="AAF22641.1"
/db_xref="GI:6653209"
/translation="MAGSPLLCGPRAGGVGILVLLGLLRLPPTLSARPVKEPRSL
AASAPLVETSTPLRLRAVRPGEAGAVQELARALAHLEAEQERARAQAEQDQ
ARVLAQLLRAGSPRASDPPLAPDDPDPAQAOLARALLRLDPAALAAQLVPAPAA
APRPVPVYDDGPTGPDVEDAGDETDVDPPELLRYLLGRITLTGSSEPEAPAPRLRR
SVDQDLGPEVPPENVLGLLRKLENPSQAPARLLPP"
sig_peptide 110..208
3'UTR 887..1027
polyA_signal 999..1004
polyA_site 1017
BASE COUNT 149 a 386 c 332 g 160 t
ORIGIN
Query Match 3.3%; Score 32; DB 10; Length 1027;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 49 TGGGGCAGCATGGCGGGTGCCTGCTCTG 80
|||||
Db 101 TGGGGCAGCATGGCGGGTGCCTGCTCTG 132
|||||
RESULT 12
AC098339 62396 bp DNA linear HTG 24-AUG-2002
LOCUS Rattus norvegicus clone CH230-2A16, *** SEQUENCING IN PROGRESS ***
DEFINITION AC098339
ACCESSION AC098339
VERSION AC098339.3 GI:22474728

KEYWORDS
SOURCE
ORGANISM

HTG: HTGS-PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 62396)

Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayalew,M., Banks,T.,
Barbala,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviado,N., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojokhan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shookstari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,T., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL

Direct Submission

Unpublished

2 (bases 1 to 62396)

Worley,K.C.

REFERENCE
AUTHORS

Direct Submission

Submitted (23-OCV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

TITLE
JOURNAL

Direct Submission

Worley,K.C.

Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 24, 2002 this sequence version replaced gi:21953722.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TULN
Center clone name: CH230-2A16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329
Consensus quality: 23320 bases at least Q40
Consensus quality: 26055 bases at least Q30
Consensus quality: 27202 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1088: contig of 1088 bp in length
1089 1188: gap of unknown length
1189 2332: contig of 1144 bp in length
2333 2432: gap of unknown length
2433 3637: contig of 1205 bp in length
3638 3737: gap of unknown length
3738 4884: contig of 1147 bp in length
4885 4984: gap of unknown length
4985 6231: contig of 1247 bp in length
6232 6331: gap of unknown length
6332 7397: contig of 1086 bp in length
7398 7498: gap of unknown length
7498 8602: contig of 1105 bp in length
8603 8702: gap of unknown length
8703 9708: contig of 1006 bp in length
9709 9808: gap of unknown length
9809 11095: contig of 1287 bp in length
11096 11195: gap of unknown length
11196 12589: contig of 1394 bp in length
12590 13867: gap of unknown length
13868 13967: contig of 1178 bp in length
13968 15150: contig of 1183 bp in length
15151 16391: contig of 1141 bp in length
16392 16491: gap of unknown length
16492 17528: contig of 1037 bp in length
17529 17628: gap of unknown length
17629 18906: contig of 1278 bp in length
18907 19006: gap of unknown length
19007 20152: contig of 1146 bp in length
20153 20252: gap of unknown length
20253 21292: contig of 1040 bp in length
21293 21392: gap of unknown length
21393 22700: contig of 1308 bp in length
22701 22800: gap of unknown length
22801 24706: contig of 1906 bp in length
24707 24806: gap of unknown length
24807 26388: contig of 1582 bp in length
26389 26488: gap of unknown length
26489 27637: contig of 1149 bp in length
27638 29102: contig of 1365 bp in length
29103 29202: gap of unknown length
29203 30352: contig of 1150 bp in length
30353 30452: gap of unknown length
30453 31720: contig of 1268 bp in length
31721 31820: gap of unknown length
31821 33903: contig of 2083 bp in length
33904 34003: gap of unknown length
34004 35759: contig of 1756 bp in length
35760 35859: gap of unknown length
35860 37240: contig of 1381 bp in length
37241 37340: gap of unknown length
37341 39222: contig of 1882 bp in length
39223 39322: gap of unknown length
39323 40923: contig of 1601 bp in length
40924 41023: gap of unknown length

```

* 41024 42882: contig of 1859 bp in length
* 42883 42982: gap of unknown length
* 42983 44647: contig of 1665 bp in length
* 44648 44747: gap of unknown length
* 44748 46546: contig of 1799 bp in length
* 46547 46646: gap of unknown length
* 46647 48461: contig of 1815 bp in length
* 48462 51223: contig of 2662 bp in length
* 51224 51323: gap of unknown length
* 51324 53891: contig of 2568 bp in length
* 53892 53991: gap of unknown length
* 53992 56351: contig of 2360 bp in length
* 56352 56451: gap of unknown length
* 56452 58948: contig of 2497 bp in length
* 58949 59048: gap of unknown length
* 59049 62396: contig of 3348 bp in length.
FEATURES             Location/Qualifiers
     source            1..62396
                        /organism="Rattus norvegicus"
                        /db_xref="taxon:10116"
                        /clone="CH230-2A16"
BASE COUNT            15738 a 13739 c 13230 g 15967 t 3722 others
ORIGIN
Query Match          3.3%; Score 32; DB 2; Length 62396;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 GCGGTGAGAGTGGCGCGCGCGCTGGCGCA 299
      |||||||||||||||||||||||||||||||
Db 54260 GCGGTGAGAGTGGCGCGCGCGCTGGCGCA 54291

RESULT 13
AC098473
LOCUS               Rattus norvegicus clone CH230-2u23, *** SEQUENCING IN PROGRESS ***,
DEFINITION          36 unordered pieces.
ACCESSION            AC098473.3 GI:21953422
VERSION              HTG; HTGS_PHASE1.
KEYWORDS             Norway rat.
SOURCE               Rattus norvegicus
ORGANISM             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                    Rattus.
REFERENCE            1 (bases 1 to 77876)
AUTHORS              Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
                    Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
                    Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
                    Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
                    Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
                    Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
                    Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,
                    Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
                    Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
                    Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
                    Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
                    Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,
                    Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
                    Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
                    Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
                    Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
                    Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
                    Jacobson,B., Jia,L., Johnson,R., Jolivet,S., Joudah,S.,
                    Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
                    Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
                    Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lousseged,H.,
                    Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
                    Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
                    Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,

```

```

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenko,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villaion,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 77876)
Worley,K.C.
Direct Submission
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 77876)
Worley,K.C.
Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 24, 2002 this sequence version replaced gl:20976551.
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUVZ
Center clone name: CH230-2J23
----- Summary Statistics
Sequencing vector: plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 42926 bases at least Q40
Consensus quality: 42726 bases at least Q30
Consensus quality: 50269 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1471: contig of 1471 bp in length
* 1472 1571: gap of unknown length
* 1572 3224: contig of 1653 bp in length
* 3225 3324: gap of unknown length
* 3325 4401: contig of 1077 bp in length
* 4402 4501: gap of unknown length
* 4502 5530: contig of 1029 bp in length
* 5531 5630: gap of unknown length
* 5631 6686: contig of 1056 bp in length
* 6687 6786: gap of unknown length
* 6787 8420: contig of 1634 bp in length
* 8421 8520: gap of unknown length
* 8521 9650: contig of 1130 bp in length
* 9651 10758: gap of unknown length
* 10759 10758: contig of 1008 bp in length
* 10759 12472: contig of 1614 bp in length

```

```

* 12473 12572: gap of unknown length
* 12573 13767: contig of 1195 bp in length
* 13768 13867: gap of unknown length
* 13868 15438: contig of 1571 bp in length
* 15439 15538: gap of unknown length
* 15539 16997: contig of 1459 bp in length
* 16998 17097: gap of unknown length
* 17098 18756: contig of 1659 bp in length
* 18757 18856: gap of unknown length
* 18857 20149: contig of 1293 bp in length
* 20150 20249: gap of unknown length
* 20250 22457: contig of 2208 bp in length
* 22458 22557: gap of unknown length
* 22558 24293: contig of 1736 bp in length
* 24294 24393: gap of unknown length
* 24394 26264: contig of 1871 bp in length
* 26265 26364: gap of unknown length
* 26365 28263: contig of 1899 bp in length
* 28264 28363: gap of unknown length
* 28364 30529: contig of 2166 bp in length
* 30530 30629: gap of unknown length
* 30630 33390: contig of 2761 bp in length
* 33391 33490: gap of unknown length
* 33491 35353: contig of 1863 bp in length
* 35354 35453: gap of unknown length
* 35454 37555: contig of 2112 bp in length
* 37556 37665: gap of unknown length
* 37666 39106: contig of 1441 bp in length
* 39107 39206: gap of unknown length
* 39207 41043: contig of 1837 bp in length
* 41044 41143: gap of unknown length
* 41144 42556: contig of 1313 bp in length
* 42557 42556: gap of unknown length
* 42557 46801: contig of 4245 bp in length
* 46802 46901: gap of unknown length
* 46902 50115: contig of 3214 bp in length
* 50116 50215: gap of unknown length
* 50216 52326: contig of 2311 bp in length
* 52327 52626: gap of unknown length
* 52627 55467: contig of 2841 bp in length
* 55468 55567: gap of unknown length
* 55568 57415: contig of 1847 bp in length
* 57416 57514: gap of unknown length
* 57515 60035: contig of 2521 bp in length
* 60036 60135: gap of unknown length
* 60136 63450: contig of 3315 bp in length
* 63451 63550: gap of unknown length
* 63551 66633: contig of 3083 bp in length
* 66634 66733: gap of unknown length
* 66734 70948: contig of 4215 bp in length
* 70949 71048: gap of unknown length
* 71049 74326: contig of 3278 bp in length
* 74327 74426: gap of unknown length
* 74427 77876: contig of 3450 bp in length.

FEATURES
    Location/Qualifiers
        1..77876
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            /clone="CH230-2J23"
BASE COUNT 18873 a 17600 c 17320 g 19098 t 4985 others
ORIGIN

```

Query Match 3.3%; Score 32; DB 2: Length 77876;
 Best Local Similarity 100.0%; Pred. No. 0.00017;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 GCGGTGAGAGCTGGCGGGCGCTGGCGCA 299

Db 38507 GCGGTGAGAGCTGGCGGGCGCTGGCGCA 38538

RESULT 14
 AC094563/c

```

LOCUS AC094563 111948 bp DNA linear RTG 11-JUL-2002
DEFINITION Rattus norvegicus clone CH230-4K5, *** SEQUENCING IN PROGRESS ***,
53 unordered pieces.
ACCESSION AC094563
VERSION AC094563.3 GI:21717879
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 111948)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayelle,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Fallis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,f., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogum,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Roife,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 111948)
Worley,K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 111948)
Worley,K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17941324.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project information

```

Center project name: GAXL
Center clone name: CH230-4K5
----- Summary Statistics -----
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 54807 bases at least Q40
Consensus quality: 60236 bases at least Q30
Consensus quality: 64720 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hscrc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1001: contig of 1001 bp in length
1101: gap of unknown length
1102: contig of 1231 bp in length
2332: gap of unknown length
2333: contig of 1053 bp in length
2433: gap of unknown length
3485: gap of unknown length
3586: contig of 1229 bp in length
4814: gap of unknown length
4914: contig of 1240 bp in length
6154: contig of unknown length
6254: gap of unknown length
7917: contig of 1663 bp in length
8017: gap of unknown length
9057: contig of 1040 bp in length
9157: gap of unknown length
10364: contig of 1207 bp in length
10365: gap of unknown length
10465: contig of 1433 bp in length
11898: gap of unknown length
11998: contig of 1266 bp in length
13264: gap of unknown length
13364: contig of 1181 bp in length
14544: gap of unknown length
14644: contig of 1420 bp in length
16064: gap of unknown length
16164: contig of 1621 bp in length
16165: gap of unknown length
17785: gap of unknown length
17886: contig of 1238 bp in length
19123: gap of unknown length
19224: contig of 1408 bp in length
20632: gap of unknown length
22214: contig of 1483 bp in length
22314: gap of unknown length
23749: contig of 1435 bp in length
23849: gap of unknown length
25423: contig of 1574 bp in length
25523: gap of unknown length
26850: contig of 1327 bp in length
26851: gap of unknown length
28487: contig of 1537 bp in length
28587: gap of unknown length
29830: contig of 1249 bp in length
29936: gap of unknown length
31672: contig of 1736 bp in length
31772: gap of unknown length
33161: contig of 1389 bp in length
33261: gap of unknown length
34437: contig of 1176 bp in length
34537: gap of unknown length
36538: contig of 2215 bp in length
36753: gap of unknown length
37909: contig of 1057 bp in length
38009: gap of unknown length
39170: contig of 1161 bp in length

39171: gap of unknown length
39271: contig of 1770 bp in length
41040: gap of unknown length
41140: contig of 1301 bp in length
42441: gap of unknown length
42542: contig of 1069 bp in length
43611: gap of unknown length
43710: contig of 1309 bp in length
45019: gap of unknown length
45119: contig of 1239 bp in length
46358: gap of unknown length
46458: contig of 2137 bp in length
48595: gap of unknown length
50326: contig of 1631 bp in length
50426: gap of unknown length
51957: contig of 1531 bp in length
52057: gap of unknown length
53961: contig of 1904 bp in length
54061: gap of unknown length
57027: contig of 2966 bp in length
57127: gap of unknown length
58760: contig of 1633 bp in length
58860: gap of unknown length
61388: contig of 2528 bp in length
61488: gap of unknown length
63296: contig of 1808 bp in length
63396: gap of unknown length
65907: contig of 2511 bp in length
66007: gap of unknown length
69562: contig of 3555 bp in length
69662: gap of unknown length
72572: contig of 2910 bp in length
72672: gap of unknown length
75269: contig of 2597 bp in length
75369: gap of unknown length
78347: contig of 2978 bp in length
78447: gap of unknown length
82487: contig of 4040 bp in length
82587: gap of unknown length
84648: contig of 2061 bp in length
84748: gap of unknown length
87601: contig of 2853 bp in length
87701: gap of unknown length
91380: contig of 3679 bp in length
91481: gap of unknown length
94959: contig of 3479 bp in length
95059: gap of unknown length
99432: contig of 4373 bp in length
99532: gap of unknown length
105198: contig of 5666 bp in length
105298: gap of unknown length
111948: contig of 6650 bp in length.

FEATURES

source
1. 111948
Query Match 3.3%; Score 32; DB 2; Length 111948;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 TGGGGCAGCATGGCGGGTCCCGCTGCTCTG 80
|||||
Db 109507 TGGGGCAGCATGGCGGGTCCCGCTGCTCTG 109476

RESULT 15

AL670169
LOCUS
DEFINITION
AL670169
ACCSSION
VERSION
KEYWORDS
house mouse.
AL670169
Mouse DNA sequence from clone RP23-198C2 on chromosome X, complete
sequence.
AL670169.8
GI:21213366
HTG.

ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 151694)
 Direct Submission
 Submitted (23-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk
 On May 25, 2002 this sequence version replaced gi:21104114.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:,
 SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-198C2 is
 from the RPCI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: PBACe3.6.

FEATURES
 source
 1..151694
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-198C2"
 /clone_lib="RPCI-23"
 BASE COUNT 39029 a 35150 c 35380 g 42135 t
 ORIGIN

Query Match 3.3%; Score 32; DB 10; Length 151694;
 Best Local Similarity 100.0%; Pred. No. 0.00016;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 49 TGGGCGAGCATGGCGGGTGGCGGCTGCTCTG 80
 ||||||||||||||||||||||||||||||||
 Db 21997 TGGGCGAGCATGGCGGGTGGCGGCTGCTCTG 22028

Search completed: April 12, 2003, 20:24:33
 Job time : 3039 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:26:00 ; Search time 45 Seconds
(without alignments)
555.443 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 260
Sequence: 1 MAGSPLLWGPRAGGVGLLVL.....RVKRLETPAPQVPARRLLPP 260

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	%
No.	Score	Match Length DB ID Description
-----	-----	-----

No matches found

Search completed: April 4, 2003, 14:29:59
Job time : 46 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:18:45 ; Search time 25 Seconds
(without alignments)
431.354 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 260
Sequence: 1 MACSPLLWGPRAGGVCLLV.....RVKRLTTPAQVPARRLLPP 260

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 112892 seqs, 41476328 residues

Word size : 10
Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: April 4, 2003, 14:27:08
Job time : 25 secs

RESULT 1
Q9UHG2
ID Q9UHG2 PRELIMINARY; PRT; 260 AA.
AC Q9UHG2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DD 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

00 SEQUENCE 230 FM, 2/203 MM, 0/424400E0000CF0 CR004,

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:26:46 ; Search time 28 Seconds
(without alignments)
273.213 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 260
Sequence: 1 MAGSPLWGPAGGVLVL.....RVKRLTPAPQVPARLLPP 260

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 10

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: April 4, 2003, 14:30:34
Job time : 28 secs


```

RESULT 1
US-09-803-589-6
; Sequence 6, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES EN
; TITLE OF INVENTION: PROGNOSTIC, D

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 17:01:47 ; Search time 1577 Seconds

(without alignments)
10054.142 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattggcagcaggccagc.....gaggatccctaccctccctgac 979

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST.*
- 1: em_estba.*
 - 2: em_esthum.*
 - 3: em_estin.*
 - 4: em_estnu.*
 - 5: em_estov.*
 - 6: em_estpl.*
 - 7: em_estro.*
 - 8: em_hc.*
 - 9: gb_est1.*
 - 10: gb_est2.*
 - 11: gb_hc.*
 - 12: gb_est3.*
 - 13: gb_est4.*
 - 14: gb_est5.*
 - 15: em_estfun.*
 - 16: em_estom.*
 - 17: gb_gss.*
 - 18: gb_gss_hum.*
 - 19: em_gss_inv.*
 - 20: em_gss_pin.*
 - 21: em_gss_vrt.*
 - 22: em_gss_fun.*
 - 23: em_gss_mam.*
 - 24: em_gss_mus.*
 - 25: em_gss_other.*
 - 26: em_gss_pro.*
 - 27: em_gss_rod.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	757	77.3	839	BI546579	BI546579 603191602
2	750.2	76.6	878	BI603988	BI603988 603244588
c	739.6	75.5	951	BI489332	BI489332 603021593
4	736.2	75.2	887	BI602351	BI602351 603251134
5	735.6	75.1	764	BI669134	BI669134 603295564
6	735.2	75.1	868	BI458350	BI458350 603198733

7 727.4 74.3 908 13 BI668879
8 724.6 74.0 745 12 BG707647
9 721 73.6 861 12 BG709387
10 712.8 72.8 817 13 BI668984
c 11 697 71.2 990 12 BF966686
12 696.6 71.2 784 13 BI668730
13 691.8 70.7 707 14 BM715636
14 688.6 70.3 771 13 BI667137
15 686.6 70.1 763 12 BG699405
16 678.8 69.3 898 13 BI544506
17 677.6 69.2 843 13 BI599557
18 677.6 69.2 883 13 BI549788
19 676 69.1 877 13 BI666575
20 670.8 68.5 735 13 BI159920
21 666 68.0 870 12 BF969262
22 663.8 67.8 740 13 BI602131
23 661.2 67.5 690 12 BG716263
c 24 658.8 67.3 692 13 BI490760
c 25 657.2 67.1 713 14 BM74078
26 649.4 66.3 857 12 BG708547
27 642 65.6 924 12 BE798423
28 624.4 63.8 681 12 BG706595
29 621.8 63.5 953 13 BI161379
30 620.6 63.4 763 13 BI602706
31 611 62.4 676 13 BI548089
32 608.8 62.2 693 13 BI603114
33 593.4 60.6 600 13 BM129679
34 589.4 60.2 736 12 BG703846
35 587.6 60.0 661 12 BG714685
36 586.8 59.9 718 10 BE383172
c 37 584 59.7 593 13 BM504127
38 581.8 59.4 586 13 BM129592
39 580.8 59.3 944 11 AK002969
40 578.2 59.1 631 12 BG108317
41 569.6 58.2 577 13 BM129728
42 569.6 58.2 585 13 BM128297
43 568.4 58.1 633 12 BG703203
44 564.8 57.7 958 12 BF310234
45 563.6 57.6 575 13 BI547813

ALIGNMENTS

RESULT 1
BI546579
LOCUS BI546579 839 bp mRNA linear EST 05-SEP-2001
DEFINITION 603191602F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262672 5', mRNA sequence.
ACCESSION BI546579
VERSION BI546579.1 GI:15433891
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 839)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Plate: L1AM1162 row: d column: 01
High quality sequence stop: 793.
Location/Qualifiers


```
Db 200 GTCTCCGCGCTTGGCTGAGACTGGCGTCTCTGCGCGTTCGCGGGGTCAGTGCCTCCCGAGG 259
QY 255 TGAGGCGCGGGGGCGGTGCAGGAGTGGCGCGCGCGCTGCGCATCTGCTGAGGCGCGA 314
Db 260 TGAGGCGCGGGGGCGGTGCAGGAGTGGCGCGCGCTGCGCATCTGCTGAGGCGCGA 319
QY 315 ACGTACAGGAGCGGGCGGGCGAGGCGCAGAGGCTGAGGATCAGAGCGGCGGCTCCT 374
Db 320 ACCTCAGGAGCGGGCGGGCGAGGCGCAGAGGCTGAGGATCAGAGCGGCGGCTCCT 379
QY 375 GCGCGACCTGCTGCGCTCTGGGCGCGCCCGCCGCACTCTGATCCGCTCTGGCTTGA 434
Db 380 GCGCGACCTGCTGCGCTCTGGGCGCGCCCGCCGCACTCTGATCCGCTCTGGGCTTGA 439
QY 435 CGAGACCCCGAGCGCGCTCAGCGCAGCTGCTGCGCTCTGCTGCGCGCGCGCTTGA 494
Db 440 CGAGACCCCGAGCGCGCTCAGCGCAGCTGCTGCGCTCTGCTGCGCGCGCGCTTGA 499
QY 495 CCCTGCGCGCTTACGACGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGGCTCCGACC 554
Db 500 CCCTGCGCGCTTACGACGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGGCTCCGACC 558
QY 555 CCGCGCGCGCTTACGACGCCAGCGCGCGCGCGGCTGCTGAGGAGCGAGCGACGA 614
Db 559 CCGCGCGCGCTTACGACGCCAGCGCGCGCGCGGCTGCTGAGGAGCGAGCGACGA 618
QY 615 GACACCCGAGCTGGACCCCGAGCTGTTGAGTACTTGTGGGACGGATTCTTGGGGAAG 674
Db 619 GACACCCGAGCTGGACCCCGAGCTGTTGAGTACTTGTGGGACGGATTCTTGGGGAAG 678
QY 675 CCGGAGCTCGAGGGGTGGACGCCCGCGCGCGCTCCCGCGCGGCTCCCGCGCGGATG 734
Db 679 CCGGAGCTCGAGGGGTGGACGCCCGCGCGCGCTCCCGCGCGGCTCCCGCGCGGATG 738
QY 735 GGGCTCTGAGCTGCGCCCTGAGGCGCTGCTGGGCGCGCTGCTGGGCGCTGCTGCTG 794
Db 739 GGGCTCTGAGCTGCGCCCTGAGGCGCGCTGCTGGGCGCGCTGCTGGGCGCTGCTGCTG 797
QY 795 GACCCCGCGCGCGAGCTGCTGACCGCGCGCTTGTCCACCTTGACACTGCGCGGATC 854
Db 798 GACCCCGCGCGCGCGCGCGCGCGCGCTTGTCCACCTTGACACTGCGCGGATC 851
QY 855 CCGTGCACCTGGGACCCA 873
Db 852 CCGTGCACCTGGGACCCA 870

RESULT 3
BI489332/c
LOCUS 603021593T1 NTH_MGC_114 Homo sapiens cDNA clone IMAGE:5192183 3',
DEFINITION mRNA sequence.
ACCESSION BI489332
VERSION BI489332.1 GI:15328560
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 951)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrived by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11480 row: j column: 24
```

```
High quality sequence start: 25
High quality sequence stop: 824.
Location/Qualifiers
1. .951
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5192183"
/lab_host="NIH_MGC_114"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."
BASE COUNT 134 a 335 c 369 g 113 t
ORIGIN
Query Match 75.5%; Score 739.6; DB 13; Length 951;
Best Local Similarity 94.1%; Pred. No. 8.8e-114;
Matches 834; Conservative 0; Mismatches 44; Indels 8; Gaps 6;
QY 75 GCTCTGGGGCGCGCGCGCGGCGCTGCGCTTGTGGTGTGCTGCTGCTGCTGCTGCTGCTT 134
Db 885 GCTGCTGGGGCGCGCGCGCGGCGCTGCGCTTGTGGTGTGCTGCTGCTGCTGCTGCTGCTT 826
QY 135 TCGGCGCGCGCGCGCGCTGCTGCGCGCGCGCGGTAAAGAGAGCGCGCGCGCTTAAGCGAGC 194
Db 825 CGGCGCGCGCGCGCGCTGCTGCGCGCGCGGTAAAGAGAGCGCGCGCGCTTAAGCGAGC 766
QY 195 GTCTCCGCGCTTGGCTGAGACTGCGCTCTGCGCGCTTCCGCGCTTCCGCGCGCGCGAG 253
Db 765 GTCTCCGCGCTTGGCTGAGACTGCGCTCTGCGCGCTTCCGCGCGCTTCCGCGCGCGAG 706
QY 254 GTGAGCGCGCGGGCGGTGCAGAGAGTGGCGC---GGGCGCTGGCGCATCTGCTGAGG 310
Db 705 GTGAAGCGCGGGCGGTGCAGGAAGTGGCGCGCGGCGAGCTGGCGCATCTGCTGAGG 646
QY 311 CCGNACCTCAGGAGCGCGCGCGCGCGCGCGAGGCGAGGAGCTGAGGATCAGAGCGCGCG 370
Db 645 CCGNACCTCAGGAGCGCGCGCGCGCGCGAGGCGAGGAGCTGAGGATCAGAGCGCGCG 586
QY 371 TCCTGGCGCAGCTGCTGCGCTCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 429
Db 585 TCCTGGCGCAGCTGCTGCGCTCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 526
QY 430 TTGAGCAGACCCCGAGCGCGCTGCGAGCGAGCTGCTGCGCTGCTGCTGCTGCTGCTGCG 489
Db 525 CTGGAGCAGACCCCGAGCGCGCTGCGAGCGAGCTGCTGCGCTGCTGCTGCTGCTGCTGCG 466
QY 490 CTTGACCTTGGCGCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 549
Db 465 CTTGACCTTGGCGCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 406
QY 550 CGACCCCGCGCGCGCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 609
Db 405 CGACCCCGCGCGCGCTTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 346
QY 610 GAGCAGACCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
Db 345 GACGAGACCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 286
QY 670 GGAAGCGCGAGCTCCGAGGCGGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
Db 285 GGAAGCGCGAGCTCCGAGGCGGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 226
QY 730 GATGTGGCTCTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789
Db 225 GATGTGGCTCTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 166
```

QY 790 CTAGAGACCCC-GGCGCCCCAGGTGCTGCACGCGCCCTCTTGCACCTGTAGCACTGCC 848
 Db 165 CTAGAGACCCCAGGCGCCAGGTCCTGCACGCGCCCTCTTGCACCTGTAGCACTGCC 106
 QY 849 CGGATCCCGTGCACCTCGGGACCCAGAGTGCCTCCGCGCATCCGCGACCAAGACTGCTC 908
 Db 105 CGGATCCCGTGA-CCTGGGACCCAGAGTGCCTCCGCGCATCCGCGACCAAGACTGCTC 47
 QY 909 CCGCGCAGACGCTCCAGAGCACTTACCCCGCGCCACGCCAGCCCTCT 954
 Db 46 CCGCGCAGACGCTCCAGAGCAAC-TAAACCGCGCGCCAGCCCTCT 2

RESULT 4
 BI602351
 LOCUS 603251134F1 NIH_MGC_96 887 bp mRNA linear EST 07-SEP-2001
 DEFINITION 603251134F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5302616 5',
 mRNA sequence.
 ACCESSION BI602351
 VERSION BI602351.1 GI:15495303
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 887)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L2AM11766 row: d column: 09
 High quality sequence stop: 764.
 Location/Qualifiers
 1. 887
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5302616"
 /clone_lib="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.3 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 94 a 334 c 336 g 123 t

BASE COUNT 94 a 334 c 336 g 123 t
 ORIGIN

Query Match 75.2%; Score 736.2; DB 13; Length 887;
 Best Local Similarity 93.4%; Pred. No. 3.2e-113;
 Matches 819; Conservative 5; Mismatches 43; Indels 10; Gaps 5;

QY 15 GCCAGCAGTCCGCGCGYMCRRGCCCGGCTCGTGGGACGATGGCGGGTCCCGCT 74
 Db 11 GCCAGCAGTCCGCGCGTCCGAGACCCCGGCTCGTGGGACGATGGCGGGTCCCGCT 70
 QY 75 GCTCTGGGGCGCGGGCGGGCGGCGTCCGCGCTTTGTGCTGCTGCTGCTGCTGCT 134
 Db 71 GCTCTGGGGCGCGGGCGGGCGGCGTCCGCGCTTTGTGCTGCTGCTGCTGCTGCT 130

QY 135 TCGGCGCCCCCGCGCTCTGCGCGGGCGCGGTAAAGGAGAGCCCGCGGCTTAAGCGCAGC 194
 Db 131 TCGGCGCCCCCGCGCTCTGCGCGGGCGGTAAAGGAGAGCCCGCGGCTTAAGCGCAGC 190
 QY 195 GTCTCCGCGCTTGGCTGAGACTGGGCTCTCTCGCGCTTCCGCGGTCAGTGCCTCCGAGG 254
 Db 191 GTCTCCGCGCTTGGCTGAGACTGGGCTCTCTCGCGCTTCCGCGGTCAGTGCCTCCGAGG 250
 QY 255 TAGAGCGCGGGGGGGTGCAGAGAGTGGCGGGGCGCTGGCGGCATCTGCTGAGAGCCGA 314
 Db 251 TAGAGCGCGGGGGGGTGCAGAGAGTGGCGGGGCGCTGGCGGCATCTGCTGAGAGCCGA 310
 QY 315 ACCTCAGAGAGCGGGCGGGCGAGCGAGAGGCTGAGGATCAGCAGCGCGGCTCT 374
 Db 311 ACCTCAGAGAGCGGGCGGGCGAGCGAGAGGCTGAGGATCAGCAGCGCGGCTCT 370
 QY 375 GCGCAGAGTGTGCGCGTCTGGGCGCGCCCGCCCAACTGATCGGCTCTCGGCTTGA 434
 Db 371 GCGCAGAGTGTGCGCGTCTGGGCGCGCCCGCCCAACTGATCGGCTCTCGGCTTGA 430
 QY 435 CGACGACCCCGACGCGCTGCAGCGAGCTCGCTCGCGCTCTGCTCCGCGCGGCTTGA 494
 Db 431 CGACGACCCCGACGCGCTGCAGCGAGCTCGCTCGCGCTCTGCTCCGCGCGGCTTGA 490
 QY 495 CCCTCCGCGCTTAGCAGCGAGCTTGTCCCGCGCGCGCTCCCGCGCGGCTCCGAGC 554
 Db 491 CCCTCCGCGCTTAGCAGCGAGCTTGTCCCGCGCGCGCTCCCGCGCGGCTCCGAGC 550
 QY 555 CCGG-CCCGCGGTCTACGAGCAGCGCGCGCGCGCGGCTGAGGAGGAGGAGCGAGC 613
 Db 551 CCGGCGCGCGGTCTACGAGCAGCGCGCGCGCGCGGCTGAGGAGGAGGAGCGAGC 610
 QY 614 AGACACCCGACGTGACCCCGCGAGCTGTTGAGGTACTTGTCTGGG--ACGGATTCTTTGCGGG 671
 Db 611 AGACACCCGACGTGACCCCGCGAGCTGTTGAGGTACTTGTCTGGGAGCGGCTTTGCGGG 670
 QY 672 --AAGCGGGACTCCGAGGGGTGCGAGCGCGCGCGCGCTCCGCGGCTCCGCGCGAGCAC 729
 Db 671 GAAGCGCGGAGTCCGAGGGGTGCGAGCGCGCGCGCGCGCTCCG-CGTGCCCGCGAGCAC 729
 QY 730 GATGTGGGCTCTGAGCTGCCCCCTGAGGCGGTCTGGGGGCGCTGCTGC---GTGTGAA 785
 Db 730 GATGTGGGCTCTGAGCTGCCCCCTGAGGCGGTCTGGGGGCGGTGCTGCTGGGCTGTA 789
 QY 786 AGCGCTAGAGACCCCGCGCGCGCGAGTGCCTGACCGCGCTCTGCGCAGCTGAGCACT 845
 Db 790 AGCGCTAGAGACCCCGCGCGCGCGAGTGCCTGACCGCGGTCTGCGCGCGCTCTGCGCGCGCT 849
 QY 846 GCCCGGATCCCGTGCACCTGGGACCCAGAGTGCCT 882
 Db 850 TGCCCGGATCCCGTGCACCTGGGACCCAGAGTGCCT 886

RESULT 5
 BI669134
 LOCUS 603295564F1 NIH_MGC_96 764 bp mRNA linear EST 12-SEP-2001
 DEFINITION 603295564F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5314814 5',
 mRNA sequence.
 ACCESSION BI669134
 VERSION BI669134.1 GI:15583367
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 764)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM11797 row: p column: 15
High quality sequence stop: 760.
Location/Qualifiers
1. .764
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5314814"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 76 a 286 c 294 g 108 t
ORIGIN
Query Match 75.1%; Score 735.6; DB 13; Length 764;
Best Local Similarity 98.7%; Pred. No. 4.1e-113;
Matches 746; Conservative 5; Mismatches 4; Indels 1; Gaps 1;
QY 15 GCACGACGTCGCCGSGYMGRCGCCGCTCGCTGGGAGCAGATGCGGGGTGCGCGCT 74
DB 10 GCCAGCAGTCGCCGCGTCCGGAGCCCGCTCGCTGGGAGCAGATGCGGGGTGCGCGCT 69
QY 75 GCTCTGGGGCCCGGGCGGGGGCTGCGGCTTTTGGTCTGCTGCTGCTGCGCGCTGTT 134
DB 70 GCTCTGGGGCCCGGGCGGGGGCTGCGGCTTTTGGTCTGCTGCTGCTGCGCGCTGTT 129
QY 135 TCGGCGCGCCCCCGCTCTGCGCGCGCGGCTAAAGAGACCCCGCGGCTAAAGCGCAGC 194
DB 130 TCGGCGCGCCCCCGCTCTGCGCGCGCGGCTAAAGAGACCCCGCGGCTAAAGCGCAGC 189
QY 195 GTCTCGCCCTTGGCTGAGACTGCGGCTCTCGCGGCTTCCGGGCTGAGTCCCGCAGG 254
DB 190 GTCTCGCCCTTGGCTGAGACTGCGGCTCTCGCGGCTTCCGGGCTGAGTCCCGCAGG 249
QY 255 TGAGGGCGGGGGCGGTGCGAGGCTGCGCGGGCGCTGCGGCATCTGCTGGAGGCCGA 314
DB 250 TGAGGGCGGGGGCGGTGCGAGGCTGCGCGGGCGCTGCGGCATCTGCTGGAGGCCGA 309
QY 315 ACCTCAGGAGCGGGCGGGCGGAGGAGGCTGAGGATCAGCAGCGCGCGCTCT 374
DB 310 ACCTCAGGAGCGGGCGGGCGGAGGAGGCTGAGGATCAGCAGCGCGCTCT 369
QY 375 GCGCAGCTCTCGGCTTGGGGCGGGCGGGCGGAGTCTGATCGGCTTGGGCTTGA 434
DB 370 GCGCAGCTCTCGGCTTGGGGCGGGCGGGCGGAGTCTGATCGGCTTGGGCTTGA 429
QY 435 CGACGACCCCGAGCGCTGACGAGCTGCTGCGGCTCTGCTGCGCGCGCGCTCGAC 494
DB 430 CGACGACCCCGAGCGGCTGACGAGCTGCTGCGGCTCTGCTGCGCGCGCGCTTGA 489
QY 495 CCCTGCGGCTTACGAGCCCGAGCTTGTCCCGCGCGCGCTGCGCGCGCGCTCGAC 554
DB 490 CCCTGCGGCTTACGAGCCCGAGCTTGTCCCGCGCGCGCTGCGCGCGCGCTCGAC 549
QY 555 CCGGCCCCCGGTACGAGACGCGGGCGGGCGGGAGTCTGAGAGGACGAGCGACGA 614
DB 550 CCGGCCCCCGGTACGAGACGCGGGCGGGCGGGAGTCTGAGAGGACGAGCGACGA 609
QY 615 GACACCGAGCTGGACCCCGAGCTTTGAGGTACTTGTGGGAGCGGATCTTCGCGGAAG 674

|||||
Db 610 GACATCGAGCTGACCCGAGCTGTGAGGTACTTCTCGGACGATTCTTCGGGAAG 669
QY 675 CGCGGACTCCGAGGGGTGGCAGCCCGCGCTCCGCGTCCGCGCCGACACCATGT 734
Db 670 CGCGGACTCCGAGGGGTGGCAGCCCGCGCTCCGCGTCCGCGCCGACACCATGT 729
QY 735 GGCTCTGAGCTGCCCTCGAGGGCGGTGCTGGGGG 770
Db 730 GGCTCTGAGCTG-CCCTGAGGGCGTGTGGGGG 764
RESULT 6
BI458350 868 bp mRNA linear EST 21-AUG-2001
LOCUS 603198733F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5277993 5',
DEFINITION mRNA sequence.
ACCESSION BI458350
VERSION BI458350.1 GI:15249019
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 868)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM11702 row: b column: 10
High quality sequence stop: 838.
Location/Qualifiers
1. .868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5277993"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 94 a 324 c 325 g 125 t
ORIGIN
Query Match 75.1%; Score 735.2; DB 13; Length 868;
Best Local Similarity 94.3%; Pred. No. 4.8e-113;
Matches 822; Conservative 5; Mismatches 38; Indels 7; Gaps 6;
QY 13 AGCCACGACGTCGCCGSGYMGRCGCCGCTCGCTGGGAGCAGATGCGGGGTGCGCG 72
Db 1 AGCGGAGTCACTCGCGCGCTCCGAGCCCGCTAGCTGGGAGCAGATGCGGGGTGCGCG 60
QY 73 CTGCTCTGGGGCGGGCGGGCGGTGCGCTTTTGGTGTGCTGCTGCTGCGCGCTG 132
Db 61 CTGCTCTGGGGCGGGCGGGCGGTGCGCTTTTGGTGTGCTGCTGCTGCGCGCTG 120
QY 133 TTTTCGCGCGCCCCCGCGGTCTGCGCGCGCGCGTAAAGGAGCCCGCGGCTTAAGCGCA 192

27	GC	SGYMCGRRRCCCGCTCGCTGGGGCAGCATGGCGGGGTGCGCGTGTGCTTGGGGGCC	86
Db	11	:::	61
	2	CGCGGTGCGAGAGCCGCTCGCTGGGGCAGCATGGCGGGGTGCGCGTGTGCTTGGGGGCC	
QY	87	CGGGCCGGGGGCGTGGCGCTTTTGGTGTGCTGTCTCGGCGCTGTTTGGCGCGCCCC	146
Db	62		121
	62	CGGGCCGGGGGCGTGGCGCTTTTGGTGTGCTGTCTCGGCGCTGTTTGGCGCGCCCC	
QY	147	CGGCTCTGCGCGGGCCGGTAAAGAGAGCCCGCGGCTTAAAGCGACGCTCTCGCCCTT	206
Db	122		181
	122	CGGCTCTGCGCGGGCCGGTAAAGAGAGCCCGCGGCTTAAAGCGACGCTCTCGCCCTT	
QY	207	GGCTGAGACTGGCGTCTCTCGCGCTTCCGCGGTAGTGCCTGAGTGGCGGCGGG	266
Db	182	GGCTGAGACTGGCGTCTCTCGCGCTTCCGCGGTAGTGCCTGAGTGGCGGCGGG	241
	267	GGCGGTGACGAGCTGGCGGGCGCTGGCGCATCTGCTGGAGGCCGAAGCTCAGGAGCG	326
Db	242	GGCGGTGACGAGCTGGCGGGCGCTGGCGCATCTGCTGGAGGCCGAACGTGAGGAGCG	301
QY	327	GGCGGGGCGAGCGCAGAGGCTGAGGATCAGCAGGCGCGCTCTGGCGGACGTGCT	386
Db	302	GGCGGGGCGAGCGCAGAGGCTGAGGATCAGCAGGCGCGCTCTGGCGGACGTGCT	361
	387	CGCGTCTGGGGCGCCCCCGCAACTCTGATCGCGTCTGGGCTTGGACGACGACCCGA	446
Db	362	CGCGTCTGGGGCGCCCCCGCAACTCTGATCGCGTCTGGGCTTGGACGACGACCCGA	421
QY	447	CGGCTGACAGCAGCTCGCTCGCGCTTGTCTCCGCGCCCGCTTGAACCTCGCGCCCT	506
Db	422	CGGCGCTGACAGCAGCTCGCTCGCGCTTGTCTCCGCGCCCGCTTGAACCTCGCGCCCT	481
	507	AGCAGCCAGCTTGTCCCGCGCCCGTCCCGCGCGGGCGCTCCGACCCCGCGCCCGGT	566
QY	482	CGCAGCCAGCTTGTCCCGCGCCCGTCCCGCGCGGCGCTCCGACCCCGG-ACCCGGT	540
Db	567	CTACGACGACGGCCCGGGCCCGATGCTGAGGAGGACAGCGACGAGACACCCAGCT	626
	541	CTACGACGACGGCCCGGGCCCGATGCTGAGGAGGACAGCGACGAGACACCCAGAT	600
QY	627	GGACCCCGAGCTTTCAGGTACTTGTGGACGGATTTTGGGGGAAGCCGGACTCCGA	686
Db	601	GGACCCCGAGCTTTCAGGTACTTGTGGACGGATTTTGGGGGAAGCCGGACTCCGA	660

Query Match	72.8%	Score 712.8	DB 13	Length 817
Best Local Similarity	96.9%	Pred. No. 2.5e-109		
Matches 785	Conservative 5	Mismatches 12	Indels 8	Gaps 6
QY	22	AGTCCGCCGSGYMGRRGCCCGCGCTCTCTGGGGAGCATGCGGGGGTCCGCGTCTCTGG	81	
Db	6	AGTCTGCCGCTCGGAGCCCGGCTCGCTGGGACAGATGCGGGGTCCGCGTCTCTGG	65	


```

QY 652 CTGGACGAGATTCTTGGCGGAAGCCGGACTCCGAGGGGGTGGCAGCCCCCGCGCCCTC 711
|||||
Db 312 CTGGACGAGATTCTTGGCGGAAGCCGGACTCCGAGGGGGTGGCAGCCCCCGCGCCCTC 253
|||||
QY 712 CGCCGTGCCCGCCGACACGATGTGGCTCTGAGCTGCCCTTGAGGGGGTGTGGGGGG 771
|||||
Db 252 CGCCGTGCCCGCCGACACGATGTGGCTCTGAGCTGCCCTTGAGGGGGTGTGGGGGG 193
|||||
QY 772 CTGCTGCGGTGTAAACGCTAGAGACCCCGGCGCCAGGTGCCCTGACGCGCCCTCTTG 831
|||||
Db 192 CTGCTGCGGTGTAAACGCTAGAGACCCCGGCGCCAGGTGCCCTGACGCGCCCTCTTG 133
|||||
QY 832 CCACCTTGAGCACTCCCGGATCCCGTGCACCTGGGACCCAGAGTGCCTCCCGCATCC 891
|||||
Db 132 CCA-CTTGAGCACTCCCGGATCCCGTGCA-CTGGGACCCAGAGTGCCTCCCGCATCC 75
|||||
QY 892 CGCCACCAAGGACTGCTCCCGCCGACGACGTCAGAGCACTTACCCCGGCGCAGCCGCC 951
|||||
Db 74 CGCCACCAAGGACTGCTCCCGCCGACGACGTCAGAGCACTTACCCCGGCGCAGCCGCC 15
|||||
QY 952 TCTCACC 958
|||||
Db 14 TCTCACC 8

RESULT 12
BI668730
LOCUS 784 bp mRNA linear EST 12-SEP-2001
DEFINITION 603293168F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5312320 5',
mRNA sequence.
ACCESSION BI668730
VERSION BI668730.1 GI:15582963
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 784)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rstraus@nihs.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11791 row: h column: 17
High quality sequence stop: 782.
FEATURES
Location/Qualifiers
1..784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5312320"
/clone.lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 81 a 300 c 296 g 107 t
ORIGIN

```

```

Query Match 71.2%; Score 696.6; DB 13; Length 784;
Best Local Similarity 97.7%; Pred. No. 1.2e-106;
Matches 722; Conservative 5; Mismatches 9; Indels 3; Gaps 2;

QY 15 GCCAGCCAGTCCCGCGGCGGRRGCCGCTCGCTGGGCGCAGATGCGGGGGTGCAGCGGT 74
|||||
Db 49 GCCAGCCAGTCCCGCGGCGGRRGCCGCTCGCTGGGCGCAGATGCGGGGGTGCAGCGGT 108
|||||
QY 75 GCTCTGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 134
|||||
Db 109 GCTCTGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 168
|||||
QY 135 TCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGGCTAAAGAGAGCGCCCGCGCTAAGCGCAGC 194
|||||
Db 169 TCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGGCTAAAGAGAGCGCCCGCGCTAAGCGCAGC 228
|||||
QY 195 GTCTCCGCGCGCTTGGCTGAGACTGGCGCTCTCTCCGCGCTTCCGCGCGGTCACTAGTCCCGCAGG 254
|||||
Db 229 GTCTCCGCGCGCTTGGCTGAGACTGGCGCTCTCTCCGCGCTTCCGCGCGGTCACTAGTCCCGCAGG 288
|||||
QY 255 TGAGCGGCGGCGGCGGCGGCTGCGAGAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 314
|||||
Db 289 TGAGCGGCGGCGGCGGCGGCTGCGAGAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 348
|||||
QY 315 AGCTCAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 374
|||||
Db 349 AGCTCAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 408
|||||
QY 375 GCGCAGCTGTGCGCGCTGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 434
|||||
Db 409 GCGCAGCTGTGCGCGCTGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 468
|||||
QY 435 CGACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 494
|||||
Db 469 CGACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 528
|||||
QY 495 CCTTCCGCGCGCTTAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 554
|||||
Db 529 CCTTCCGCGCGCTTAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 588
|||||
QY 555 CGCGCGCGCGGCTTACGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 614
|||||
Db 589 CGCGCGCGCGGCTTACGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 648
|||||
QY 615 GACACCGCGGCTGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 674
|||||
Db 649 GACACCGCGGCTGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 708
|||||
QY 675 CGCGGACTCCGAGGGGGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 734
|||||
Db 709 CGCGGACTCCGAGGGGGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 765
|||||
QY 735 GGGCTCTCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 753
|||||
Db 766 GGGCTCTCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 784
|||||

RESULT 13
BI715636
LOCUS 707 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-EJ0-ahl-1-14-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone
UI-E-EJ0-ahl-1-14-0-UI 5', mRNA sequence.
ACCESSION BI715636
VERSION BI715636.1 GI:19028894
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 707)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene

```

discovery
 JOURNAL
 MEDLINE
 COMMENT
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 256-326, >CC-rich#Low_complexity
 Seq primer: M13 Reverse.

FEATURES

source
 Location/Qualifiers
 1..707
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ0-ahi-1-14-0-UI"
 /clone_lib="UI-E-EJ0"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina Foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-EJ0 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes, AGAATCAAGA
 ; lens, CGATTACCGA; eye anterior segment, AATGCCGATC;
 optic nerve, CCATTAAATG; retina, CCGCG; Retina Foveal and
 Macular, GTC; RPE and Choroid, ACCTA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."
 BASE COUNT 91 a 285 c 240 g 90 t
 ORIGIN

Query Match 70.7%; Score 691.8; DB 14; Length 707;
 Best Local Similarity 99.4%; Pred. No. 7.9e-106;
 Matches 704; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
 QY 267 GCGGTGACAGGAGCTGCGCGGCGCTGGCGATCTGCTGGAGCGGACGTCAGGAGCG 326
 Db 1 GCGGTGACAGGAGCTGCGCGGCGCTGGCGATCTGCTGGAGCGGACGTCAGGAGCG 60
 QY 327 GCGCGGGCGGAGCGCAGGAGCTGAGGATCAGCAGGCGCGCTGCTGGCGAGCTGCT 386
 Db 61 GCGCGGGCGGAGCGCAGGAGCTGAGGATCAGCAGGCGCGCTGCTGGCGAGCTGCT 120
 QY 387 GCGCGTCTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 446
 Db 121 GCGCGTCTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 QY 447 GCGCGTCTGAGCGACGCTGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 506
 Db 181 GCGCGTCTGAGCGACGCTGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 240
 QY 507 AGCAGCCGAGCTGTGTCGCGCGCGCGCTGCCCGCGCGCGCGCTCCGACCGCGCGCGG 566
 Db 241 AGCAGCCGAGCTGTGTCGCGCGCGCGCTGCCCGCGCGCGCGCTCCGACCGCGCGCGG 300

QY 567 CTACGACGAGCGCGCGCGCGCGCGGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 626
 Db 301 CTACGACGAGCGCGCGCGCGCGCGGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
 QY 627 GGACCCCGAGCTGCTGAGGTACTTGTGGACGAGATCTTGGGGAAGCGCGGAGCTCCGA 686
 Db 361 GGACCCCGAGCTGCTGAGGTACTTGTGGACGAGATCTTGGGGAAGCGCGGAGCTCCGA 420
 QY 687 GGGGTGGCAGCGCGCGCGCGCGCTCCGCGCGTCCGCGGACGAGATGTGGGCTCTGAGCT 746
 Db 421 GGGGTGGCAGCGCGCGCGCGCGCTCCGCGCGTCCGCGGACGAGATGTGGGCTCTGAGCT 480
 QY 747 GCGGCTGAGGGGCTGCTGGGGGCGCTGCTGTGTGAAACGCTAGAGACCGCGCGCC 806
 Db 481 GCGGCTGAGGGGCTGCTGGGGGCGCTGCTGTGTGAAACGCTAGAGACCGCGCGCC 540
 QY 807 CAGGTGCTGACGCGCGCGCTCTTGCACCTGAGCAGCTGCCGGATCCCGTGCACCCCTG 866
 Db 541 CAGGTGCTGACGCGCGCGCTCTTGCACCTGAGCAGCTGCCGGATCCCGTGCACCCCTG 600
 QY 867 GGACCCAGAGTGCCTCCCGCGCGGATCCCGCGGAGGAGTGTCCCGCGGAGGAGGAGGAG 926
 Db 601 GGACCCAGAGTGCCTCCCGCGCGGATCCCGCGGAGGAGTGTCCCGCGGAGGAGGAGGAG 660
 QY 927 GCACTTACCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 974
 Db 661 GC-ACCTACCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 707

RESULT 14
 BI667137
 LOCUS
 DEFINITION BI667137 771 bp mRNA linear EST 12-SEP-2001
 603291530F1 NTH_MGC_96 Homo sapiens cDNA clone IMAGE:5310870 5',
 mRNA sequence.
 ACCESSION BI667137
 VERSION BI667137.1 GI:15581370
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 771)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM1787 row: 1 column: 07
 High quality sequence stop: 763.

FEATURES

source
 Location/Qualifiers
 1..771
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5310870"
 /clone_lib="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.3 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library.
BASE COUNT 78 a 296 c 292 g 105 t

Query Match 70.3%; Score 688.6; DB 13; Length 771;
Best Local Similarity 97.7%; Pred. No. 2.7e-105; Indels 3; Gaps 3;
Matches 724; Conservative 5; Mismatches 9;

Qy 15 GCCAGCAGTCCGCGCGYMCRRGCCGCTCGTGGGCGACATGCGGGGTCGCCGCT 74
|||||
Db 30 GCCAGCAGTCCGCGCGTCCGAGGCGCGCTCGTGGGCGACATGCGGGGTCGCCGCT 89
Qy 75 GCTCTGGGGGGCGCGGGCGGGCGCTCGGCGCTTTTGTGCTGCTGCTGCTGCTGCTGCT 134
Db 90 GCTCTGGGGGGCGCGGGCGGGCGCTCGGCGCTTTTGTGCTGCTGCTGCTGCTGCTGCT 149

Qy 135 TCGGGCGCGCGCGCGCTCGGCGCGCGCGGTAAAGGAGCGCGCGCGCTAAAGCGCAGC 194
Db 150 TCGGGCGCGCGCGCGCTCGGCGCGCGCGGTAAAGGAGCGCGCGCGCGCTAAAGCGCAGC 209

Qy 195 GTCTCCGCGCTTGGCTGAGACTGGCGCTTCTCGCGCTTCCGGCGGTCAAGTCCCGCAGG 254
Db 210 GTCTCCGCGCTTGGCTGAGACTGGCGCTTCTCGCGCTTCCGGCGGTCAAGTCCCGCAGG 269

Qy 255 TGAGCGCGGGGGGGGTGACAGAGCTGGCGCGGGCGCTGCGCGCATCTGCTGGAGCGCA 314
Db 270 TGAGCGCGGGGGGGGTGACAGAGCTGGCGCGGGCGCTGCGCGCATCTGCTGGAGCGCA 329

Qy 315 ACCTCAGAGCGGGCGCGCGCGGAGCGCGAGGCTCAGATCAGCAGGCGCGGCTCT 374
Db 330 ACCTCAGAGCGGGCGCGCGCGGAGCGCGAGGCTCAGATCAGCAGGCGCGGCTCT 389

Qy 375 GCGCAGCTGTGCGCTGTGGGGCGCGCGCGCGCAACTCTGATCGGCTCTGGGCTTGA 434
Db 390 GCGCAGCTGTGCGCTGTGGGGCGCGCGCGCGCAACTCTGATCGGCTCTGGGCTTGA 449

Qy 435 CGACAGCGCGCGCGCGCTGACGCGCAGCTCGTGGCGCTCTGCTCCGCGCGCGCTTGA 494
Db 450 CGACAGCGCGCGCGCGCTGACGCGCAGCTCGTGGCGCTCTGCTCCGCGCGCGCTTGA 509

Qy 495 CCCTCGCGCGCTAGCAGCGCGCGCTGTGCGCGCGCGCGCTTCCGCGCGCGCTCCGACC 554
Db 510 CCCTCGCGCGCTAGCAGCGCGCGCTGTGCGCGCGCGCGCTTCCGCGCGCGCTCCGACC 569

Qy 555 CCGCGCGCGCTTACGACGCGCGCGCGCGCGCGGATGCTGAGGAGGCGGCGAGCA 614
Db 570 CCGCGCGCGCTTACGACGCGCGCGCGCGCGGATGCTGAGGAGGCGGCGAGCA 629

Qy 615 GACACCGG - ACCTGACCGCGCGCTGTGAGGTACTTCTGCGAGCGGATCTTGGGGGAA 673
Db 630 GACACCGGAGCTGACCGCGCGCGCTGTGAGGTACTTCTGCGAGCGGATCTTGGGGGAA 689

Qy 674 GCGGCGACTCCGA - GGGGGTGGACCGCGCGCGCGCTTCCGCGCGCGCGCGACCAT 732
Db 690 GCGGCGACTCCGAAGGGGGTGGACCGCGCGCGCGCTTCCGCGCGCGCGCGACCAT 749

Qy 733 GTGGG - CTCTGAGCTGCCCG 752
Db 750 GTGGGCTCTGAGCTGGCGCG 770

RESULT 15
BG699405
LOCUS
DEFINITION 602679088F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4812140 5',
763 bp mRNA linear EST 07-MAY-2001
mRNA sequence.
ACCESSION BG699405
VERSION BG699405.1 GI:13967665
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 763)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki,
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10703 row: o column: 21
High quality sequence stop: 763.

FEATURES
source
1..763
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4812140"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTT-3', size-selected for average insert size 2.5 kb and normalized to 50% of total. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 80 a 280 c 297 g 106 t

Query Match 70.1%; Score 686.6; DB 12; Length 763;
Best Local Similarity 96.4%; Pred. No. 5.7e-105;
Matches 728; Conservative 5; Mismatches 19; Indels 3; Gaps 3;

Qy 15 GCCAGCAGTCCGCGCGYMCRRGCCGCTCGTGGGCGACATGCGGGGTCGCCGCT 74
Db 12 GCCAGCAGTCCGCGCGTCCGAGGCGCGCTCGTGGGCGACATGCGGGGTCGCCGCT 71

Qy 75 GCTCTGGGGGGCGCGGGCGGGCGCTGCGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCT 134
Db 72 GCTCTGGGGGGCGCGGGCGGGCGCTGCGCTTTTGTGCTGCTGCTGCTGCTGCTGCT 131

Qy 135 TCGGCGCGCGCGCGCGCTTCTCGCGCGCGCGGTAAAGGAGCGCGCGCGCTAAAGCGCAGC 194
Db 132 TCGGCGCGCGCGCGCGCTTCTCGCGCGCGCGGTAAAGGAGCGCGCGCGCTAAAGCGCAGC 191

Qy 195 GTCTCCGCGCTTGGCTGAGACTGGCGCTTCTCGCGCGCTTCCGCGCGGTCAAGTCCCGCAGG 254
Db 192 GTCTCCGCGCTTGGCTGAGACTGGCGCTTCTCGCGCGCTTCCGCGCGGTCAAGTCCCGCAGG 251

Qy 255 TGAGCGCGGGGGCGGGTGGAGGAGCTGGCGCGCGCGCTGGCGCATCTGCTGGAGGCGCA 314
Db 252 TGAGCGCGGGGGCGGGTGGAGGAGCTGGCGCGCGCGCTGGCGCATCTGCTGGAGGCGCA 311

Qy 315 ACCTCAGAGCGCGCGCGCGCGGAGGCGCGAGGCTGAGGATCAGCAGGCGCGGCTCT 374
Db 312 ACCTCAGAGCGCGCGCGCGCGGAGGCGCGAGGCTGAGGATCAGCAGGCGCGGCTCT 371

Qy 375 GCGCAGCTGCTGCGCGCTGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
Db 372 GCGCAGCTGCTGCGCGCTGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 431

Qy 435 CGACGACCGCGCGCGCTGCGAGCGAGCTGCTGCGCGCTGCTGCTGCGCGCGCGCGCGCGCG 494
Db 432 CGACGACCGCGCGCGCTGCGAGCGAGCTGCTGCGCGCTGCTGCTGCGCGCGCGCGCGCGCG 491

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	77.8	7.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C 2	71.6	7.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
3	69.6	7.1	888	4	US-08-765-907A-2	Sequence 1, Appli
C 4	69.6	7.1	2888	4	US-08-765-907A-1	Sequence 1, Appli
5	69.4	7.1	4411529	4	US-09-103-840A-1	Sequence 4, Appli
6	62.6	6.4	936	1	US-08-018-977C-4	Sequence 1, Appli
C 7	60.4	6.2	4411529	4	US-09-103-840A-1	Sequence 1, Appli
8	59.6	6.1	4488	4	US-08-406-030A-3	Sequence 3, Appli
9	59.4	6.1	4524	2	US-08-843-998-7	Sequence 7, Appli
10	59.4	6.1	4524	3	US-09-206-537-7	Sequence 7, Appli
11	59.4	6.1	4524	4	US-09-430-854-7	Sequence 7, Appli
12	59.2	6.0	4257	2	US-08-690-473-1	Sequence 1, Appli
13	59.2	6.0	4257	4	US-09-253-821A-1	Sequence 1, Appli
14	59.2	6.0	4257	4	US-08-843-659-1	Sequence 1, Appli
C 15	59.2	6.0	12001	1	US-08-458-568A-11	Sequence 11, Appl
16	58.8	6.0	8438	1	US-07-945-283-1	Sequence 1, Appli
C 17	58.4	6.0	2823	1	US-08-398-008A-1	Sequence 1, Appli
C 18	58.4	6.0	2823	2	US-08-893-333-1	Sequence 1, Appli
C 19	58.2	5.9	30001	1	US-08-123-468-1	Sequence 1, Appli
C 20	58.2	5.9	30001	2	US-08-474-933-1	Sequence 1, Appli
21	58	5.9	44377	2	US-08-804-227C-7	Sequence 7, Appli
22	58	5.9	44377	2	US-08-804-198-1	Sequence 1, Appli
C 23	57.8	5.9	4257	2	US-08-690-473-1	Sequence 1, Appli
C 24	57.8	5.9	4257	4	US-09-253-821A-1	Sequence 1, Appli
C 25	57.8	5.9	4257	4	US-08-843-659-1	Sequence 1, Appli
C 26	57.8	5.9	12001	1	US-08-458-568A-11	Sequence 11, Appl
C 27	57.6	5.9	1030	3	US-08-853-003-2	Sequence 2, Appli

Qy 336 CGAGCGCAGAGGCTGAGGATCAGACGGCGGCGCTCTCTGGCGCAGCTGCTCGCGGTCTG 395

Db 337937 CCGCGCGCGGCCACCGCGGTTCGCCGCGCGCCGCGAAGACCGCGCGCCG -GGCGCGC 337995

Qy 3396 GGGCGCCCCCGCAACTCTGATCCGGCTCTGGGCTTGGACGACGACCCGCGACGCGCTGC 455

Db 337996 GCGACCGCGCCCGCGCTGAGACACCGGCTTTCGCGCGGCGCGCGCGCCCGCGCG 338055

Qy 456 AGCGCAGCTCGCTCGCGCTCTGCTCCGCGCGCGCTTGACCTGTCGCGCTTAGCAGCCCA 515

Db 338056 CGTCAAGCCACCGCCGCGCAGCGCGCGGGCGCCCGCGAGCCGAACAGTGTCCGCGCT 338115

Qy 516 GCTTGTCCCCGCGCGCTCCCGCGGCGCGCTCCCGACCCGCGCCCCCGTCTACGACGA 575

Db 338116 CGCTCCGATCCACCGCGCAGCGCGCCACCGTCCGGCTGGATCCGCGTGC CGCGG 338175

Qy 576 CGGCCCCCGGGCGCGGATGTGAGGAGGCGAGCGACGAGACACCGCAGTGGACCCCGA 635

Db 338176 CGCGCGCGGCGCACCGAGGCTGAGCATGCGGGCTCGCGCGCGCCACCGCTTCCGCG 338235

Qy 636 GCTGTTGAGGTACTTGCTGGGACGGATCTTGGCGGAAGCGGACTCCGAGGGGTGCG 695

Db 338236 CGAGTTTGATTTAGTCTCGTCCGCGCACTACCGCGGTCCGCGCGCGCGACACCGCGC 338295

Qy 696 AGCCCGCGCGCGCTCGCCGTGCGCGACACGATGCGCTGAGCTGCGCCCTGA 755

Db 338296 CAGAGCCGCCATCCCGCGCGCGCGCGCCCAAGATCCGAAATCCGCGGGGCCACCGG 338355

Qy 756 GGGCGTGTGGGGCGCTGCTCGGTGTAAACGCTACAGACCCCGCGCGCCAGGTGCC 815

Db 338356 TGCC--GCGCGCGCGACACCGCGCGCTTTCGCGCGGATCCGCGGGCCCCCGCGTGC 338413

Qy 816 TGC 818

Db 338414 GGC 338416

RESULTS

```

RESUL 2
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent NO. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R. M.
APPLICANT: FRASER, John C.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007-00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

	Query Match	7.3%	Score 71.6;	DB 4;	Length 4403765;
	Best Local Similarity	44.9%	Pred. No. 0.00016;		
	Matches 361;	Conservative	0;	Mismatches 434;	Indels 9; Gaps 2;
QY	120	GCTGCTGGGCTGTTTGGCGCGCCCGCGCGCTCTGCGCGCGCGGTAAAGGAGCCCCG	179		
Db	3929274	GCTGTGTTTCCGTTCTGGCGCGTACACCGGCGCGCGGTGCGCGCGTGGCGCGGCG	3929215		
QY	180	CGGCTTAAGCGCAGCGTCTCGCGCCTTGGTGTAGACTGGCGCTCTCGCGCGCTTCCGCGC	239		

[illegible]

RESULT 3

RESULT 3
US-08-765-907A-2
Sequence 2, Application us/08765907A
Patent No. 6352839
GENERAL INFORMATION:
APPLICANT: BLANC, Veronique
APPLICANT: THIBAUT, Denis
APPLICANT: BAMA-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSSCHE, Laurent
APPLICANT: FAMEUCHON, Alain
APPLICANT: PARIS, Jean-Marc
APPLICANT: DUTRUC-ROSSET, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
TITLE OF INVENTION: Mutasyntesis
FILE REFERENCE: Streptogramin genes
CURRENT APPLICATION NUMBER: US/08/765,907A
CURRENT FILING DATE: 1997-03-20
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0


```
QY 681 CTCGAGGGGGTGCAGCCCGCGCGCTCCGCGTGGCGCGGACACGATGTGGCTC 740
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 GGGGCGGGGCTGGAGGGTGGGACGGGCGCTCCCGCGGCGCGCCCTCCGCGCCCG 734
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 741 TGAGCTGCCCTGAGGGGCTGTGGGGGCGTGTGCTGTGAACGCTAGAGACCC 800
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 735 CCGGCTCCCTCCGCGCTCTCTCGAGCAGGAGGCGGCTCGCTCCGAGGAGA- 790
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 801 GGGGCCCCAGGTGCTGCAGCGCGCTCTTGCCACCTGAGCACTGCGCGGATCCCGTGC 860
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 791 ---GGCGCGGGTGGGCCAGGAAACCCCTCCCGCAGTCCAGCGCTCCCGCCCTCGCGC 847
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 861 ACCTGGGACCCAGAGTGCCTCCCGCATCCCGCACACGAGGACTGTCTCCCGCCACG 920
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 848 CGCGAGGGCCAGAGGGGCGGACGACACCCCGCTCCGACTCAGGGCGCGGGGGCGCG 907
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 921 TCAGAG 927
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 908 GCCAGG 914
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6254328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 6.2%; Score 60.4; DB 4; Length 4411529;
Best Local Similarity 46.4%; Pred. No. 0.011;
Matches 365; Conservative 0; Mismatches 408; Indels 14; Gaps 4;

QY 191 CAGCGTCTCGGCTGGCTGAGACTGGCGCTCTCGCGCTTCCGCGGTGAGTCCGCC 250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3936062 CGGCGCGGCTCTGCTGCCAGCGCTGTGCTTCCGCTTCCGCGGTGAGTCCGCC 3936003

QY 251 GAGGTGAGCGGGGGGGGCTGCAGAGCTGGCGGGGCGTGGCGCATCTGCTGGAGG 310
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3936002 CGGTCCCGCGGTGCGCGCGCGCGCTGTGATGCCCGCGCGCGGTGCGCGGACCC 3935943

QY 311 CCGAAGCTCAGAGCGGGCGCGGGCGCGAGAGCTGAGGATCAGCAGCGCGCG 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935942 CGCCTTGGCGGCTTGGCGCGCGGGAAGCGCTTGCCTTGGGAGAGGGCGCGGT 3935883

QY 371 TCTTGGCGAGTGTGCGGCTGTGGGGGCGCGCGCGCACTGATCGGCTGTGGGT 430
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935882 CGCGGCGCGCGCGCTGC-----CGCGCGCGCGCGCGCGCGATACCGGCGTTGCC 3935827

QY 431 TGAGCAGACACCGCGCGCGCTGCAGCGAGCTGCTCGCGCTGTGCTCGCGCGCGCC 490
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935826 GGGTTCGCGCGGTGTGGCCAGCGCGCGCGAGCGCGCGGTGCGCGCGCTCCGCGGTG 3935767

QY 491 TTGACCTGCGCGCTTACGAGCGCGAGCTTGTCCCGCGCGCGCTCCCGCGCGCGCTCC 550
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935766 CGCGCGGACCGCGCGCAAGAGCGCTGCGCTGCGCGCGGTGGCGCGCGCGCGCGCC 3935707

QY 551 GACCGCGCGCGCGCTGTACGAGAGCGCGCGCGCGCGCGCGGATGTGAGGAGCGAGCGG 610
```

```
Db 3935706 CCGCGCTTGCCTCCCGCTACCGCTGGCGGCGTGGCTAGCCCGGTGCTGTACGGCGCGCT 3935647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 611 ACGAGACACCCGACGACGACCCCGAGCTGTGAGGTACTTGTGGGACGAGTTCCTTTCGG 670
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935646 TTG-----CCGCTTGGCGCGCTGGCGCGGACACCTTGGCGCGCTCAATGCTG 3935593
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 671 GAAGCCGCGACTCCGAGGGGTGGACGCCCGCGCGCTTCCGCCCTGCGCGCGCGACG 730
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935592 GGGTCCCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935533
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 731 ATGTGGGCTCTGA-GCTGCCCTGAGGGGCTGTGGGGGCGTGTGCTGTGAAACGC 789
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935532 CGGTCCCGGGAAACACCGCGCGGATCCGGGTTTGGCGCGCGCGCGCGCGCGCT 3935473
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 790 CTAGAGACCCCGCGCGCGCGCTGTGACGCGCGCTTGTTCACCTGAGCACTGCC 849
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935472 TGGCGCGCTACCGCTTGGCGCGCGCGCGCGCGCTGCTGCGCACGCTGCTGTCG 3935413
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 850 GGATCCCGTGCACCTGGGACCCAGAAAGTGCCCGCGCGCTGCGCGCGCGCGCGCG 909
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935412 TTCTGGCGCTCACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935356
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 910 CGCCGAGCAGTCCAGAGCAACTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935355 CGCGCGCGCGCGCGCTTGC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 970 ACCCGCT 976
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3935295 TTGCGCT 3935289
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-08-406-030A-3/c
; Sequence 3, Application US/08406030A
; Patent No. 6270989
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Hauge, Brian M.
; APPLICANT: Selden, Richard F.
; TITLE OF INVENTION: Protein Production and Delivery
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.c.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/406,030A
; FILING DATE: 17-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/243,391
; FILING DATE: 13-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,586
; FILING DATE: 03-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/911,533
; FILING DATE: 10-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,840
; FILING DATE: 05-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/789,188
```

[illegible]

Db 3074 CCTTGGCGGCTCGGCGCGCTGCGCGCGATGCGCGCTGGATGCGCCAGATCCCGGACC 3133
RESULT 14
US-08-843-659-1
; Sequence 1, Application US/08843659
; Patent No. 6218103
; GENERAL INFORMATION:
; APPLICANT: Leopardi, Roasrio
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS US3 AND ICP4 AS
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; CITY: Houston
; STATE: Texas
; COUNTRY: United States
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARSB:519
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4257 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-843-659-1
Query Match 6.0%; Score 59.2; DB 4; Length 4257;
Best Local Similarity 44.5%; Pred. NO. 0.024;
Matches 374; Conservative 0; Mismatches 453; Indels 13; Gaps 3;
QY 48 CTGGGGCAGCATGGCGGGTGGCGGCTGCTGCGGGGCGCGGGGCGGGGGGGGGGGTGGCGCT 107
Db 2307 CGGCGGCGAGCGAGCGCGCGCGTGGCGCGCTGCGCGCGCTGAGCCTGGTGGCGGGCGCT 2366
QY 108 TTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 167
Db 2367 GGGCGCGCGCTGCGCGGGAGCGCGCGCTGCGGAGCTGCGCGCGCGCGCGCGCGCGGA 2426
QY 168 AAGAGAGCGCGCGCGCTGAAGCGCAGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCG 227
Db 2427 CCGTGTGTTTACAAACAGAGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCG 2486
QY 228 CCGCTTCGGCGGTGAGTGGCGCGAGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 287
Db 2487 GCGCGCGCGAGCGGTGGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 2542
QY 288 GCGCTGGCGATGCTGCTGGAGCGCGAAGTTCAGGAGCGCGCGCGCGCGCGCGCGCGCG 347
Db 2543 AGCGCAAGAGTCCCG 2602
QY 348 GCGTGAGGATCAGCAGCGCGCGCTGCTGGCGCAGCTGCTGCGCGCTGCTGGCGCGCGCG 407
Db 2603 CGAAGAGCGCGCGCGAGCGCGCGCGCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCG 2662

QY 408 CAACTCTGATCCGGCTCTGGGCTTGGACGACGAGACACCGCGCTTGCAGCGAGCTCGC 467
Db 2663 CGCGCGCGCTCCAGCG 2720
QY 468 TCGCGCTCTGCTCGCGCGCGCGCTTACACCTGCGCGCGCTAGCAGCGCGCTTGTCTCC 527
Db 2721 GCG 2780
QY 528 GCGCGTCCCGCGCGCGCGCTCCGACCGCGCGCGCGCGCGCGCTACGACGAGCGCGCG 587
Db 2781 CCGCGACCGCGCTGGCGGCTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2840
QY 588 CCGGATGCTGAGGAGCGAGCGAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCT 647
Db 2841 CGCGCGCGCGCGCGCGCGCTACTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 2900
QY 648 CTTGCTGGGACGATCTTGGCGGAGCGCGGACTCCGAGGCGGCTGCGAGCGCGCGCG 707
Db 2901 CCGCGTGTCCCGCTCCCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2953
QY 708 CCGCGCGCTGCGCGCGCGCGCGCGCGCTGAGCTGCGCGCGCGCGCGCGCGCGCGCG 767
Db 2954 TGGCGCTGATCG 3013
QY 768 GCGCGTGTGCTGTGAACGCGCTAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 827
Db 3014 GCGGCGAGGACGAGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3073
QY 828 CTTGCGACCGCTGAGCAGCTGCGCGGATCCGCTGCGCGCGCGCGCGCGCGCGCGCG 887
Db 3074 CCCTGCGCGCTGCGCGCGCGCGCGCGCGATGGCGCGCTGGATGGCGCGCGCGCGCG 3133
RESULT 15
US-08-458-568A-11/c
; Sequence 11, Application US/08458568A
; Patent No. 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; APPLICANT: Yeh, Lily
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
; TITLE OF INVENTION: Infections
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,568A
; FILING DATE: 02-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,146
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn R.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: DFCI-0029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12001 base pairs

Job time : 9767 secs

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Herpes simplex virus
; STRAIN: Herpes Simplex Virus Type 1
US-08-458-568A-11

Query Match      6.0%; Score 59.2; DB 1; Length 12001;
Best Local Similarity 44.5%; Pred. No. 0.023;
Matches 374; Conservative 0; Mismatches 453; Indels 13; Gaps 3;

QY 48 CTGGGGCAGCATGGCGGGTGGCGCTGCTTGGGGGCGCGGGGCGGCGGCGCTTCGGCCT 107
Db 4367 CGGCGCAGCAGGCGCGCGTGGCGCGCTGCGCGCTGAGCTGCTGCCGGGGCCT 4308

QY 108 TTTGGTCTGCTGCTGCTCGGCTGTTTCGGCGCGCGCGCGCGCTGCGCGCGCGCGGT 167
Db 4307 GGGCGCGCGCTGCGCGGACCGCGCGCTGCGGAGCTCGCGCGCGCGCGCGCGCGGA 4248

QY 168 AAGGAGCGCGCGCGCTAAGCGCAGCGTCTCGGCCCTTGGCTGAGACTGCGCGCTCCTCG 227
Db 4247 CCTGCTTTTGCACACAGAGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGGA 4188

QY 228 CCGCTTCCGGGCTGAGTGGCGCGAGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 287
Db 4187 CGCGCGCGCGCGCTGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGGAGG---GGCGCA 4132

QY 288 GCGCGTGGCGCATCTGCTGAGGCGCGGACGTCAGGAGCGCGCGCGCGCGCGCGCGAGGA 347
Db 4131 AGCGCAAGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAAGA 4072

QY 348 GGCTGAGGATCAGCAGCGCGCGCTGCTGGCGCAGCTGCTCGCGCTCTGGGGCGCGCGCGCG 407
Db 4071 CGAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4012

QY 408 CAACTCTGATCCGGCTCTGGGCTTGGACGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 467
Db 4011 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3954

QY 468 TCGCGCTCTGCTCGCGCGCGCGCTTACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 527
Db 3953 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3894

QY 528 GCCCGTCCCGCGCGCGCGCTCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 587
Db 3893 CCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3834

QY 588 CCGCGATGCTGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 647
Db 3833 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3774

QY 648 CTTGCTGGGAGGATCTTCGGGAGCGCGCGGACTCGGAGGGGTGGCGCGCGCGCGCGCG 707
Db 3773 CCGCGCTGTTCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3721

QY 708 CTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 767
Db 3720 TGGCTCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3661

QY 768 GCGCGCTGCTGCTGTGAACGCGCTAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 827
Db 3660 GCGCGCAGCAGCAGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3601

QY 828 CTTGCCACCGCTGAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 887
Db 3600 CCCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3541
```

Search completed: April 12, 2003, 21:16:56

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 18:14:02 ; Search time 151 Seconds
(without alignments)

5687.068 Million cell updates/set

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcgacagggccagc.....gaggatccctaccctccggc 979

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2.6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2.6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2.6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2.6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2.6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2.6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2.6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2.6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2.6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2.6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2.6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2.6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2.6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2.6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	977	99.8	979	9	US-09-803-589-5		Sequence 5, Appli
2	912.8	93.2	1050	9	US-09-984-271-58		Sequence 58, Appl
3	507.8	51.9	1027	10	US-09-803-589-13		Sequence 13, Appl
c	4	276.6	28.3	301	10	US-09-880-107-1701	Sequence 1701, Appl
5	274.4	28.0	349	10	US-09-960-352-71		Sequence 71, Appl
6	244.8	25.0	372	10	US-09-983-965-2663		Sequence 2663, Appl
7	167	17.1	342	10	US-09-983-965-3684		Sequence 3684, Appl
8	75.2	7.7	1614	9	US-09-976-740-45		Sequence 45, Appl
9	75.2	7.7	1614	12	US-10-023-529-45		Sequence 45, Appl
10	75.2	7.7	1614	12	US-10-023-523-45		Sequence 45, Appl
11	75.2	7.7	12425	9	US-09-976-740-50		Sequence 50, Appl
12	75.2	7.7	12425	12	US-10-023-529-50		Sequence 50, Appl
13	75.2	7.7	12425	12	US-10-023-523-50		Sequence 50, Appl
14	73.4	7.5	2561	9	US-09-976-740-48		Sequence 48, Appl
15	73.4	7.5	2561	12	US-10-023-529-48		Sequence 48, Appl
16	73.4	7.5	2561	12	US-10-023-523-48		Sequence 48, Appl
17	69.4	7.1	2307	9	US-09-893-519A-87		Sequence 87, Appl
c	18	63.2	6.5	12425	9	US-09-976-740-50	Sequence 50, Appl
c	19	63.2	6.5	12425	12	US-10-023-529-50	Sequence 50, Appl

c	20	63.2	6.5	12425	12	US-10-023-523-50	Sequence 50, Appl
c	21	62.8	6.4	12733	9	US-10-032-393-47	Sequence 47, Appl
c	22	62.8	6.4	12739	9	US-10-032-393-8	Sequence 8, Appl1
c	23	62.4	6.4	4653	9	US-10-101-388-2	Sequence 2, Appl1
c	24	62.4	6.4	8036	9	US-10-101-388-1	Sequence 1, Appl1
c	25	62	6.3	790	9	US-10-004-717-20	Sequence 20, Appl
c	26	60.4	6.2	1614	9	US-09-976-740-45	Sequence 45, Appl
c	27	60.4	6.2	1614	12	US-10-023-529-45	Sequence 45, Appl
c	28	60.4	6.2	1614	12	US-10-023-523-45	Sequence 45, Appl
c	29	59.2	6.0	4257	9	US-09-825-288A-1	Sequence 1, Appl1
c	30	58.2	5.9	2529	10	US-09-815-242-7852	Sequence 7852, Ap
c	31	58.2	5.9	37113	9	US-09-999-121-11	Sequence 11, Appl
c	32	58.2	5.9	43058	10	US-09-954-456-292	Sequence 292, App
c	33	58.2	5.9	43058	10	US-09-954-456-529	Sequence 529, App
c	34	58.2	5.9	43058	10	US-09-880-107-3950	Sequence 3950, Ap
c	35	57.8	5.9	2561	9	US-09-976-740-48	Sequence 48, Appl
c	36	57.8	5.9	2561	12	US-10-023-529-48	Sequence 48, Appl
c	37	57.8	5.9	2561	12	US-10-023-523-48	Sequence 48, Appl
c	38	57.8	5.9	4257	9	US-09-825-288A-1	Sequence 1, Appl1
c	39	57.6	5.9	1030	9	US-09-735-056-2	Sequence 2, Appl1
c	40	57.6	5.9	41936	10	US-09-967-768A-116	Sequence 116, App
c	41	57.4	5.9	88421	9	US-09-976-059-1	Sequence 1, Appl1
c	42	57	5.8	1758	10	US-09-815-242-7825	Sequence 7825, Ap
c	43	57	5.8	1817	10	US-09-924-417-64	Sequence 64, Appl
c	44	56.4	5.8	987	10	US-09-772-656-5	Sequence 5, Appl
c	45	56.4	5.8	987	10	US-09-772-656-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-803-589-5
; Sequence 5, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 979
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(837)
US-09-803-589-5

Query Match 99.8%; Score 977; DB 10; Length 979;
Best Local Similarity 100.0%; Pred. No. 1.2e-208;
Matches 979; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAATTCGCACAGAGCCAGCCAGTCCGCCSGTMCRRRCCCGGCTCGTGGGCGACATG 60
|
Db 1 GAATTCGCACAGAGCCAGCCAGTCCGCCSGTMCRRRCCCGGCTCGTGGGCGACATG 60
|
QY 61 GCGGGTCCGCGCTCTCTGGGGCCGCGGCGCGGGGCGTGGCTTTTGGTGTCTGCTG 120
|
Db 61 GCGGGTCCGCGCTCTCTGGGGCCGCGGCGCGGGGCGTGGCTTTTGGTGTCTGCTG 120
|
QY 121 CTGTCGCGCCCTGTTTGGCCGCCGCCGCTCTGCGCGCGCGCGGTAAGAGAGCCCGC 180
|
Db 121 CTGTCGCGCCCTGTTTGGCCGCCGCCGCTCTGCGCGCGCGCGGTAAGAGAGCCCGC 180
|
QY 181 GGCCTAAGCGAGGCTCTCCGCCCTTGCTGAGACTGCGCTCTCTCGCCGCTTCGGGGG 240
|
Db 181 GGCCTAAGCGAGGCTCTCCGCCCTTGCTGAGACTGCGCTCTCTCGCCGCTTCGGGGG 240
|
QY 241 TCAGTGCCTCCGAGGTAGGCGCGCGGGGCGGTGCAGGAGCTGGCGGCGCTGGCGCAT 300
|
Db 241 TCAGTGCCTCCGAGGTAGGCGCGCGGGGCGGTGCAGGAGCTGGCGGCGCTGGCGCAT 300
|
QY 301 CTGTGGAGGCGCAACGTCAGAGAGCGGCGCGCGCGAGGCGCGAGGCTGAGGATCAG 360
|
Db 301 CTGTGGAGGCGCAACGTCAGAGAGCGGCGCGCGCGAGGCGCGAGGCTGAGGATCAG 360
|
QY 361 CAGGGCGGCTCTGGGCGAGCTGTGCGCGCTGTGGGGCGCGCGCGCGCAACTCTGATCCG 420
|
Db 361 CAGGGCGGCTCTGGGCGAGCTGTGCGCGCTGTGGGGCGCGCGCGCGCAACTCTGATCCG 420
|
QY 421 GCTCTGGGCTTGGAGAGAGCCGCGAGCGCTGCAGCGCAGCTGCTGCGGCTCTGCTC 480
|
Db 421 GCTCTGGGCTTGGAGAGAGCCGCGAGCGCTGCAGCGCAGCTGCTGCGGCTCTGCTC 480
|
QY 481 CGCGCCGCTTGGAGAGAGCCGCGAGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCG 540
|
Db 481 CGCGCCGCTTGGAGAGAGCCGCGAGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCG 540
|
QY 541 CGGGGCTTCGACCGCGCGCGCGCTTACGACGAGCGCGCGCGCGCGCGCGCGCGCG 600
|
Db 541 CGGGGCTTCGACCGCGCGCGCGCTTACGACGAGCGCGCGCGCGCGCGCGCGCGCG 600
|
QY 601 GAGCAGCGAGAGACACCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
|
Db 601 GAGCAGCGAGAGACACCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
|
QY 661 ATTCTGCGGGAAGCGGAGTCCGAGGGGTGCGAGCGCGCGCGCGCGCGCGCGCGCG 720
|
Db 661 ATTCTGCGGGAAGCGGAGTCCGAGGGGTGCGAGCGCGCGCGCGCGCGCGCGCGCG 720
|
QY 721 GCCGACACGATGTGGGCTGTGAGCTGCCCGCTGAGGGCGTGTGGGGGCGCTGCTGCT 780
|
Db 721 GCCGACACGATGTGGGCTGTGAGCTGCCCGCTGAGGGCGTGTGGGGGCGCTGCTGCT 780
|
QY 781 GTGAAGCGCTTAGAGACCGCGCGCGCGAGTGCCTGCAGCGCGCGCTTGTCCACCTGA 840
|
Db 781 GTGAAGCGCTTAGAGACCGCGCGCGCGAGTGCCTGCAGCGCGCGCTTGTCCACCTGA 840
|
QY 841 GCATGCGCGGATCCCGTGCACCTTGGGAGCCAGAAAGTGCCTCCCGCGCGCGCGCG 900
|
Db 841 GCATGCGCGGATCCCGTGCACCTTGGGAGCCAGAAAGTGCCTCCCGCGCGCGCGCG 900
|
QY 901 GACTGTCCCGCGAGCAGCTTCAGAGCACTTACCCGCGCGCGCGCGCGCGCGCGCG 960
|
Db 901 GACTGTCCCGCGAGCAGCTTCAGAGCACTTACCCGCGCGCGCGCGCGCGCGCGCG 960
|
QY 961 AGGATCCCTACCCCTGGC 979
|
Db 961 AGGATCCCTACCCCTGGC 979
|
```

RESULT 2
US-09-984-271-58
; Sequence 58, Application US/09984271

```
; Publication No. US20030040088A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/984,271
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/482,273
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US99/15849
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: 60/092,921
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,922
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,956
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 1050
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-984-271-58
```

```
Query Match 93.2%; Score 912.8; DB 9; Length 1050;
Best Local Similarity 98.9%; Pred. No. 2e-194;
Matches 955; Conservative 5; Mismatches 2; Indels 4; Gaps 4;

QY 14 GGCAGCAGTCCGCCSGTMCRRRCCCGGCTCGTGGGCGAGCATGGCGGGTCCGCCG 73
|
Db 11 GGCAGCAGTCCGCCCGCTCCGAGCCCGGCTCGTGGGCGAGCATGGCGGGTCCGCCG 70
|
QY 74 TGCCTCTGGGGCCGCGGCGCGGCGTTCGCGCTTTTGGTGTCTGCTGCTCGGCTGT 133
|
Db 71 TGCCTCTGGGGCCGCGGCGCGGCGTTCGCGCTTTTGGTGTCTGCTGCTCGGCTGT 130
|
QY 134 TTCGGCCGCCCGCGCTCTGCGCGCGCGGTAAGAGAGCCCGCGGCGCTTAAGCGAG 193
|
Db 131 TTCGGCCGCCCGCGCTCTGCGCGCGCGGTAAGAGAGCCCGCGGCGCTTAAGCGAG 190
|
QY 194 CGTCTCGCCCTTGGCTGAGACTGGCGCTCTCGCGCTTCGGGGGTCACTGGCCGAG 253
|
Db 191 CGTCTCGCCCTTGGCT-AGACTGGCGCTCTCGCGCTTCGGGGGTCACTGGCCGAG 249
|
QY 254 GTGAGCGCGCGGCGGCTGCAGGAGCTGGCGCGGCGCTGGCGCATCTGCTGAGGCG 313
|
Db 250 GTGAGCGCGCGGCGGCTGCA-GACTGGCGCGGCGCTGGCGCATCTGCTGGA-GCCG 307
|
QY 314 AACCTCAGGAGCGGCGGCGGCGAGGCGAGGCTGAGGATCAGCAGCGCGCGCTCC 373
|
Db 308 AACCTCAGGAGCGGCGGCGGCGAGGCGCA-GAGGCTGAGGATCAGCAGCGCGCTCC 366
|
QY 374 TGGCGCAGCTCTCGCGCTTGGGGCGCGCGCGCACTCTGATCCGGCTTGGGCTGG 433
|
Db 367 TGGCGCAGCTCTCGCGCTTGGGGCGCGCGCGCACTCTGATCCGGCTTGGGCTGG 426
|
QY 434 ACAGCAGCCCGAGCGCGCTGCAGCAGCTGCGGCTGCTGCTCGCGCGCGCTTCG 493
|
Db 427 ACAGCAGCCCGAGCGCGCTGCAGCAGCTGCGGCTGCTGCTCGCGCGCGCTTCG 486
|
QY 494 ACCCTGCGCGCTAGCAGCCAGCTTGTCCCGCGCGCGCGCGCGCGCGCTTCGAC 553
|
Db 487 ACCCTGCGCGCTAGCAGCCAGCTTGTCCCGCGCGCGCGCGCGCGCGCTTCGAC 546
|
QY 554 CCGGGCGCGGCTTACGAGCAGCGCGCGCGCGCGCGCGCGCGCTTGGAGGAGCGAG 613
|
Db 547 CCGGGCGCGGCTTACGAGCAGCGCGCGCGCGCGCGCGCGCGCTTGGAGGAGCGAG 606
|
QY 614 AGACACCGAGCTGAGCCCGAGCTTGTGAGTACTTGTGGGAGCGGATCTTTCGGGAA 673
|
Db 607 AGACACCGAGCTGAGCCCGAGCTTGTGAGTACTTGTGGGAGCGGATCTTTCGGGAA 666
|
```


QY 790 CTAGAGACCCCGCGCCAGGTGCC 815
Db 676 CGCCCGCGCGCAGCCACAGCGCC 701

RESULT 11

US-09-976-740-50
; Sequence 50, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 12425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-740-50

Query Match 7.7%; Score 75.2; DB 9; Length 12425;
Best Local Similarity 48.6%; Pred. No. 9.5e-09;
Matches 275; Conservative 0; Mismatches 278; Indels 13; Gaps 2;
QY 261 GCGGGGGGGTGCAGGAGTGGCGGCGTGGCGCATCTGCTGAGGCGCGAAGCTCA 320
Db 2969 GCGCGCGCGGACCTGGAGCGCATCTGCGGATGGTGGCGGCGCAGCGCCGAGGC 3028
QY 321 GGAGCGGGCGCGCGCGAGCGCAGGAGGTGAGGATCAGAGCGCGCGCTCTG----- 375
Db 3029 GGAGCGCAGCGCGCGCGAGCTCGAGAACTGATCCAGCAGCGCGCTGCTCGGGTTCAG 3088
QY 376 -----GCGCAGTGTGCGCGTCTGGGGCGCCCGCGCAACTCTGATCGGCTCTGGGC 429
Db 3089 CTAAAGGGGAGCATCTCGTACCGCAACGCGCGCGCTCCAGCGCCCGCGCGGAGC 3148
QY 430 TTGAGCAGACACCCGACGCGCTGCAGGCGAGCTCGCTCGCGCTCTGCTCCGCGCCGC 489
Db 3149 CACCCCGCGGCG 3208
QY 490 CTTGACCTTCCCGCTAGAGCCAGCTTGTTCCTCCGCGCGCGCTCCCGCGCGCGCTC 549
Db 3209 GCGCGCGCGCACCG 3266
QY 550 CGACCCCG 609
Db 3267 GCG 3326
QY 610 GAGGAGACACCCGACGTGGACCCGAGCTTGTGAGGTACTTGTGGAGCGGATCTTGGC 669
Db 3327 CAGCCGGGCG 3386
QY 670 GGAAGCGCGAGCTCCGAGGGGGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
Db 3387 GCTCCCGCGCGTGGCG 3446
QY 730 GATGTGGGCTCTAGCTGTCCCGCTGAGAGGGGTGCTGGGGGCGCTGCTGTGTGAACGC 789

Db 3447 GCCCGGAGCCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3506
QY 790 CTAGAGACCCCGCGCCAGGTGCC 815
Db 3507 CGCCCGCGCGCAGCCACAGCGCC 3532

RESULT 12

US-10-023-529-50
; Sequence 50, Application US/10023529
; Patent No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 12425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-023-529-50

Query Match 7.7%; Score 75.2; DB 12; Length 12425;
Best Local Similarity 48.6%; Pred. No. 9.5e-09;
Matches 275; Conservative 0; Mismatches 278; Indels 13; Gaps 2;
QY 261 GCGGGGGGGTGCAGGAGTGGCGGCGTGGCGCATCTGCTGAGGCGCGAAGCTCA 320
Db 2969 GCGCGCGCGGACCTGGAGCGCATCTGCGGATGGTGGCGGCGCAGCGCCGAGGC 3028
QY 321 GGAGCGGGCGCGCGCGAGCGCAGGAGGTGAGGATCAGAGCGCGCGCTCTG----- 375
Db 3029 GGAGCGCAGCGCGCGCGAGCTCGAGAACTGATCCAGCAGCGCGCTGCTCGGGTTCAG 3088
QY 376 -----GCGCAGTGTGCGCGTCTGGGGCGCCCGCGCAACTCTGATCGGCTCTGGGC 429
Db 3089 CTAAAGGGGAGCATCTCGTACCGCAACGCGCGCGCTCCAGCGCCCGCGCGGAGC 3148
QY 430 TTGAGCAGACACCCGACGCGCTGCAGGCGAGCTCGCTCGCGCTCTGCTCCGCGCCGC 489
Db 3149 CACCCCGCGGCG 3208
QY 490 CTTGACCTTCCCGCTAGAGCCAGCTTGTTCCTCCGCGCGCGCTCCCGCGCGCGCTC 549
Db 3209 GCGCGCGCGCACCG 3266
QY 550 CGACCCCG 609
Db 3267 GCG 3326
QY 610 GAGGAGACACCCGACGTGGACCCGAGCTTGTGAGGTACTTGTGGAGCGGATCTTGGC 669
Db 3327 CAGCCGGGCG 3386

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: April 12, 2003, 19:29:20 ; Search time 1589 Seconds

(without alignments)

9978.214 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gatttcgacgagccagc.....gaggtacccctaccctcgtgc 979

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 15

Total number of hits satisfying chosen parameters: 95542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estc:*

9: gb_estl:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estcom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_roo:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	612	62.5	707	14	BM715636
2	561	57.3	784	13	BI668730
3	517	52.8	839	13	BI546579
4	514	52.5	764	13	BI669134
5	504	51.5	570	12	BF438085
6	487	49.7	589	14	BM876317

7	484	49.4	599	14	BM695925	BM695925	UI-E-CL1-
8	483	49.3	537	14	BM720023	BM720023	UI-E-EJ0-
c	478	48.8	593	13	BM504127	BM504127	UI-E-EJ0-
c	474	48.4	601	13	BI711080	BI711080	UI-E-EJ0-
c	472	48.2	515	13	BM052768	BM052768	UI-E-EJ0-
c	472	48.2	518	13	BI789216	BI789216	UI-E-EJ0-
c	472	48.2	541	12	BF724971	BF724971	UI-E-EJ0-
c	472	48.2	562	13	BM142527	BM142527	UI-E-EJ0-
c	472	48.2	569	13	BI789206	BI789206	UI-E-EJ0-
c	472	48.2	573	13	BI789119	BI789119	UI-E-EJ0-
c	472	48.2	589	13	BM129392	BM129392	UI-E-EJ0-
c	472	48.2	590	13	BI789105	BI789105	UI-E-EJ0-
c	472	48.2	596	13	BM129323	BM129323	UI-E-EJ0-
c	471	48.1	713	14	BM674078	BM674078	UI-E-EJ0-
c	467	47.7	619	14	BQ268412	BQ268412	UI-E-EJ0-
c	464	47.4	476	10	BE045542	BE045542	UI-E-EJ0-
c	464	47.4	498	14	BQ581370	BQ581370	UI-E-EJ0-
c	463	47.3	692	13	BI490760	BI490760	UI-E-EJ0-
c	462	47.2	631	12	BG108317	BG108317	UI-E-EJ0-
c	459	46.9	586	13	BM129592	BM129592	UI-E-EJ0-
c	454	46.4	626	13	BM508023	BM508023	UI-E-EJ0-
c	453	46.3	504	12	BF725537	BF725537	UI-E-EJ0-
c	453	46.3	581	13	BI963705	BI963705	UI-E-EJ0-
c	451	46.1	559	13	BM504353	BM504353	UI-E-EJ0-
c	450	46.0	817	13	BI668984	BI668984	UI-E-EJ0-
c	449	45.9	604	13	BM508441	BM508441	UI-E-EJ0-
c	448	45.8	617	13	BI711823	BI711823	UI-E-EJ0-
c	444	45.4	469	9	AI796185	AI796185	UI-E-EJ0-
c	444	45.4	469	9	BQ102329	BQ102329	UI-E-EJ0-
c	443	45.3	469	9	AI261620	AI261620	UI-E-EJ0-
c	442	45.1	600	13	BM129679	BM129679	UI-E-EJ0-
c	439	44.8	584	14	BQ582055	BQ582055	UI-E-EJ0-
c	434	44.3	475	13	BM510525	BM510525	UI-E-EJ0-
c	434	44.3	482	13	BM511355	BM511355	UI-E-EJ0-
c	433	44.2	771	13	BI667137	BI667137	UI-E-EJ0-
c	431	44.0	481	12	BG819901	BG819901	UI-E-EJ0-
c	431	44.0	517	13	BM510394	BM510394	UI-E-EJ0-
c	431	44.0	878	13	BI603988	BI603988	UI-E-EJ0-
c	430	43.9	690	12	BG716263	BG716263	UI-E-EJ0-

ALIGNMENTS

RESULT 1	BM715636	707 bp	mRNA	linear	EST 28-FEB-2002
LOCUS	UI-E-EJ0-ahi-1-14-0-UI-r2	UI-E-EJ0	Homo sapiens	cdna	clone
DEFINITION	UI-E-EJ0-ahi-1-14-0-UI 5', mRNA sequence.				
ACCESSION	BM715636				
VERSION	BM715636.1	GI:19028894			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 707)				
COMMENT	Bonaldo, M.F., Lennon, G. and Soares, M.B.				
	Normalization and subtraction: two approaches to facilitate gene				
	discovery				
	Genome Res. 6 (9), 791-806 (1996)				
	97044477				
	Contact: Soares, MB				
	Program for Rat Gene Discovery and Mapping				
	University of Iowa				
	451 Eckstein Medical Research Building Iowa City, IA 52242, USA				
	Tel: 319 335 8230				
	Fax: 319 335 9565				
	Email: msoares@blue.weeg.uiowa.edu				
	Tissue Procurement: Dr. Gregg Hageman				
	CDNA Library preparation: Dr. M. Bento Soares, University of Iowa				
	CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa				
	CDNA Sequencing by: Dr. M. Bento Soares, University of Iowa				


```
Db 132 CGCTCGGCTTTGGTGGCTCTGCTCGCTCGGCTGTTTCGGCGCCGCCCGCGCTTGCG 191
QY 158 CGCGGGCGGTAAAGAGAGCCCGCGGCTTAAGCGCAGCGTTCGCGCCCTTGCTGAGACTG 217
Db 192 CGCGGGCGGTAAAGAGAGCCCGCGGCTTAAGCGCAGCGTTCGCGCCCTTGCTGAGACTG 251
QY 218 GCCTCTCGCCCTTCGCGGCTAGTCCCGAGTCAAGTCCCGAGTCAAGTCCCGAGTCAAGT 277
Db 252 GCCTCTCGCCCTTCGCGGCTAGTCCCGAGTCAAGTCCCGAGTCAAGTCCCGAGTCAAGT 311
QY 278 AGCTGCGCGGCGCTGCGGCTAGTCCCGAGTCAAGTCCCGAGTCAAGTCCCGAGTCAAGT 337
Db 312 AGCTGCGCGGCGCTGCGGCTAGTCCCGAGTCAAGTCCCGAGTCAAGTCCCGAGTCAAGT 371
QY 338 AGCGCAGGAGGCTGAGGATCAGCAGCGCGGCTCTGCGGCTAGTCCCGAGTCAAGTCCCG 397
Db 372 AGCGCAGGAGGCTGAGGATCAGCAGCGCGGCTCTGCGGCTAGTCCCGAGTCAAGTCCCG 431
QY 398 CGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGGACGACGACCCGAGCGCCCTGCAG 457
Db 432 GCAGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGGACGACGACCCGAGCGCCCTGCAG 491
QY 458 CGCAGCTCGCTCGGCTCTGCTCCGCGCGGCTTGGACGACGACCCGAGCGCCCTGCAG 517
Db 492 CGCAGCTCGCTCGGCTCTGCTCCGCGCGGCTTGGACGACGACCCGAGCGCCCTGCAG 551
QY 518 TTGTCCTCCGCGCGGCTCCGCGCGGCTTGGACGACGACCCGAGCGCCCTGCAG 577
Db 552 TTGTCCTCCGCGCGGCTCCGCGCGGCTTGGACGACGACCCGAGCGCCCTGCAG 611
QY 578 GCGCGCGCGCGGCTTGGACGACGACCCGAGCGCCCTGCAG 637
Db 612 GCGCGCGCGCGGCTTGGACGACGACCCGAGCGCCCTGCAG 671
QY 638 TGTGAGGTAAGTCTGCGGAGCGATCTTTCGGGAAAGCGGAGTCCGAGGGGCTGCGAG 697
Db 672 TGTGAGGTAAGTCTGCGGAGCGATCTTTCGGGAAAGCGGAGTCCGAGGGGCTGCGAG 731
QY 698 CCC 700
Db 732 CCC 734

RESULT 3
BI546579
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 839)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@baylor.edu
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML162 row: d column: 01
High quality sequence stop: 793.
Location/Qualifiers
```

```
source
1. .839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10b"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 87 a 311 c 324 g 117 t
ORIGIN
Query Match 52.8%; Score 517; DB 13; Length 839;
Best Local Similarity 99.7%; Pred. No. 2.2e-226;
Matches 617; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 38 GCCCGGCTCGCTGGGCGAGCATGGCGGGTCCGGCTGCTCTGGGGCGCGGGCGGGG 97
Db 22 GCCCGGCTCGCTGGGCGAGCATGGCGGGTCCGGCTGCTCTGGGGCGCGGGCGGGG 81
QY 98 GCCTCGGCTTTTGGTGGCTGCTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 157
Db 82 GCCTCGGCTTTTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 141
QY 158 CGCGCGCGGTAAAGAGAGCCCGCGGCTTAAGCGCAGCGTCTCGCGCCCTTGCTGAGACTG 217
Db 142 CGCGCGCGGTAAAGAGAGCCCGCGGCTTAAGCGCAGCGTCTCGCGCCCTTGCTGAGACTG 201
QY 218 GCCTCTCGCCCTTCGCGGCTTGGGCTGAGTCCCGAGTCAAGTCCCGAGTCAAGT 277
Db 202 GCCTCTCGCCCTTCGCGGCTTGGGCTGAGTCCCGAGTCAAGTCCCGAGTCAAGT 261
QY 278 AGCTGCGCGGCGCTTGGCGCATCTGCTGAGGCGCGAAGCTCAGGAGCGCGCGCGCG 337
Db 262 AGCTGCGCGGCGCTTGGCGCATCTGCTGAGGCGCGAAGCTCAGGAGCGCGCGCGCG 321
QY 338 AGCGCAGGAGGCTGAGGATCAGCAGCGCGGCTTGGCGCAGCTGCTGCGCGCTTGGG 397
Db 322 AGCGCAGGAGGCTGAGGATCAGCAGCGCGGCTTGGCGCAGCTGCTGCGCGCTTGGG 381
QY 398 GCGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGGACGACGACCCGAGCGCCCTGCAG 457
Db 382 GCGCGCCCGCCCACTCTGATCCGGCTTGGGCTTGGACGACGACCCGAGCGCCCTGCAG 441
QY 458 CGCAGCTCGCTCGGCTCTGCTCCGCGCGGCTTGGACCTGCGCGCTTGGACGACCCGAGC 517
Db 442 CGCAGCTCGCTCGGCTCTGCTCCGCGCGGCTTGGACCTGCGCGCTTGGACGACCCGAGC 501
QY 518 TTGTCCTCCGCGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCT 577
Db 502 TTGTCCTCCGCGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCT 561
QY 578 GCGCGCGGCGCGGATGCTGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 637
Db 562 GCGCGCGGCGCGGATGCTGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 621
QY 638 TGTGAGGTAAGTCTGCTGGG 656
Db 622 TGTGAGGTAAGTCTGCTGGG 640
RESULT 4
BI669134
LOCUS
DEFINITION
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
764 bp mRNA linear EST 12-SEP-2001
BI669134
LOCUS
DEFINITION
603295564F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5314814 5',
mRNA sequence.
```

ACCESSION BI669134
VERSION BI669134.1 GI:15583367
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/
1 (bases 1 to 764)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11797 row: p column: 15
High quality sequence stop: 760.
FEATURES
source
1..764
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5314814"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 76 a 286 c 294 g 108 t
ORIGIN
Query Match 52.5%; Score 514; DB 13; Length 764;
Best Local Similarity 99.6%; Pred. No. 5.1e-225;
Matches 664; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 38 GCCGGCTCGTGGGAGCAGATGGGGGTCGGCGCTCTCTGGGGCGCGGGCGGGG 97
|||||
DB 33 GCCGGCTCGTGGGAGCAGATGGGGGTCGGCGCTCTCTGGGGCGCGGGCGGGG 92
|||||
QY 98 GCGTCGGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCG 157
|||||
DB 93 GCGTCGGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCG 152
|||||
QY 158 CGCGCGCGGTAAAGAGCGCCCGCGCTTAAGCAGCGCTCTCCGCCCTTGCTGAGACTG 217
|||||
DB 153 CGCGCGCGGTAAAGAGCGCCCGCGCTTAAGCAGCGCTCTCCGCCCTTGCTGAGACTG 212
|||||
QY 218 GCGTCTCTCGCGCTTTCGGGGGTCAGTGGCCCCGAGGTGAGCGGGCGGGCGGTGAGG 277
|||||
DB 213 GCGTCTCTCGCGCTTTCGGGGGTCAGTGGCCCCGAGGTGAGCGGGCGGGCGGTGAGG 272
|||||
QY 278 AGCTGGCGCGGCGCTGCGCATCTGCTGGAGGCGCGAACGTCAGAGCGGGCGGGCGG 337
|||||
DB 273 AGCTGGCGCGGCGCTGCGCATCTGCTGGAGGCGCGAACGTCAGAGCGGGCGGGCGG 332
|||||
QY 338 AGGCGCAGGAGGCTGAGGATCAGCAGCGCGCGCTCTGGCGAGCTGCTGGCGCTCTGG 397
|||||
DB 333 AGGCGCAGGAGGCTGAGGATCAGCAGCGCGCGCTCTGGCGAGCTGCTGGCGCTCTGG 392
|||||
QY 398 GCGCGCGCGCAACTGATCCGCTTGGGCTTGGAGGAGGAGCGCGCGCTTGAG 457
|||||

Db 393 GCGCCCCCGCAACTCTGATCCGGCTTGGGCTTGGAGGAGGAGCGCGCTTGCGAG 452
QY 458 GCGAGCTCGCTCGCGCTCTGCTCCCGCGCGCTTGGCCCTTGGCCCTTAGCAGCCGAGC 517
|||||
Db 453 GCGAGCTCGCTCGCGCTCTGCTCCCGCGCGCTTGGCCCTTGGCCCTTAGCAGCCGAGC 512
|||||
QY 518 TTGTCGCCCGCGCTCCCGCGCGCGCTCCCGACCCCGCGCGCTTACGACGAGC 577
|||||
Db 513 TTGTCGCCCGCGCTCCCGCGCGCGCTCCCGACCCCGCGCGCTTACGACGAGC 572
|||||
QY 578 GCGCGCGCGCGCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 637
|||||
Db 573 GCGCGCGCGCGCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 632
|||||
QY 638 TGTGAGGTACTTGTGCGGAGCGGATTTCTGGGGAAGCGCGGACTCCGAGGGGGTGGCAG 697
|||||
Db 633 TGTGAGGTACTTGTGCGGAGCGGATTTCTGGGGAAGCGCGGACTCCGAGGGGGTGGCAG 692
|||||
QY 698 CCGCGCG 704
|||||
Db 693 CCGCGCG 699
|||||

RESULT 5
BF438085/c
LOCUS
DEFINITION BF438085 570 bp mRNA linear EST 30-MAR-2001
7666g11.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3703485 3'
similar to TR:Q9UHG2 Q9UHG2 PROSAS PRECURSOR; contains TAR1.t3
TAR1 repetitive element; mRNA sequence.
ACCESSION BF438085
VERSION BF438085.1 GI:11450602
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 570)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 479.
Location/Qualifiers
1..570
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3703485"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73p-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and as circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 76 a 170 c 238 g 86 t
ORIGIN

```
Query Match          51.5%; Score 504; DB 12; Length 570;
Best Local Similarity 100.0%; Pred. No. 1.9e-220;
Matches 504; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 476 TGCTCCGCGCCGCTTACCTGCGCCCTAGCAGCCAGCTTGTCGCCGCGCGCTCC 535
Db 536 TGCTCCGCGCCGCTTACCTGCGCCCTAGCAGCCAGCTTGTCGCCGCGCGCTCC 477
QY 536 CGCGCCGCGGCTTCCGACCCCGCGCCCTTACGACGACGCGCCCGCGCGCGATG 595
Db 476 CGCGCCGCGGCTTCCGACCCCGCGCCCTTACGACGACGCGCCCGCGCGCGATG 417
QY 596 CTGAGGAGCAGCGGACGACACCGAGCTGGACCCCGAGCTGTGAGTACTTCTGG 655
Db 416 CTGAGGAGCAGCGGACGACACCGAGCTGGACCCCGAGCTGTGAGTACTTCTGG 357
QY 656 GACGGATTCTTGGGGAAGCGGAGCTCCGAGGGGTGGCAGCCCGCGCGCTCCGCC 715
Db 356 GACGGATTCTTGGGGAAGCGGAGCTCCGAGGGGTGGCAGCCCGCGCGCTCCGCC 297
QY 716 GTGCCGCCGACACGATGCGGCTGTGAGCTGCCCCCTGAGGCGGTGCTGGGGCGCTGC 775
Db 296 GTGCCGCCGACACGATGCGGCTGTGAGCTGCCCCCTGAGGCGGTGCTGGGGCGCTGC 237
QY 776 TGCGTGTGAACCCCTAGAGACCCCGCGCCCGAGTGCTGCACGCGCGCTTGCAC 835
Db 236 TGCGTGTGAACCCCTAGAGACCCCGCGCCCGAGTGCTGCACGCGCGCTTGCAC 177
QY 836 CTGAGCACTGCGCGGATCCGCTGCACCTGGACCCAGCAAGTGGCCCGCATCCCGC 895
Db 176 CTGAGCACTGCGCGGATCCGCTGCACCTGGACCCAGCAAGTGGCCCGCATCCCGC 117
QY 896 ACCAGACTGTCTCCCGCGCAGACGCTCCAGAGCACTTACCCTCCGCGCAGCCCTCTC 955
Db 116 ACCAGACTGTCTCCCGCGCAGACGCTCCAGAGCACTTACCCTCCGCGCAGCCCTCTC 57
QY 956 ACCGAGGATCCCTACCCCTGGC 979
Db 56 ACCGAGGATCCCTACCCCTGGC 33

RESULT 6
BM876317/c
LOCUS
DEFINITION
i161a10.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5635123
3' similar to TR:Q9UH2 Q9UH2 PROSAS PRECURSOR. ;, mRNA sequence.
BM876317
VERSION
BM876317.1 GI:19243983
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 589)
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blisstein,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other_ESTs: i161a10.y1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pbluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
```

University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -400P from Gibco
High quality sequence stop: 415.
Location/Qualifiers
1..589
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5635123"
/tissue_type="Human insulinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:
XhoI; Site:2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

BASE COUNT 78 a 173 c 237 g 100 t 1 others

Query Match 49.7%; Score 487; DB 14; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.2e-212;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 GACCTTGCCGCGCTTACGACCCAGCTTGTCGCCGCGCGCTCCCGCGCTCCGA 552
Db 529 GACCTTGCCGCGCTTACGACCCAGCTTGTCGCCGCGCGCTCCCGCGCTCCGA 470
QY 553 CCCGCGCCCGCTTACGACCCAGCTTGTCGCCGCGCGCTCCCGCGCTCCGA 612
Db 469 CCCGCGCCCGCTTACGACCCAGCTTGTCGCCGCGCGCTCCCGCGCTCCGA 410
QY 613 GAGACACCCGCTTACGACCCAGCTTGTCGCCGCGCTTGTCGCCGCGCTTGTCGCCGGA 672
Db 409 GAGACACCCGCTTACGACCCAGCTTGTCGCCGCGCTTGTCGCCGCGCTTGTCGCCGGA 350
QY 673 ACAGCGGACTCCGAGGGGTGGCAGCCCGCGCGCTCCCGCGCTCCCGCGCGCTCCCGCGGAT 732
Db 349 ACAGCGGACTCCGAGGGGTGGCAGCCCGCGCGCTCCCGCGCTCCCGCGCGCTCCCGCGGAT 290
QY 733 GTGGGCTCTGAGCTGCCCTTACGACCCAGCTTGTCGCCGCGCTTGTCGCCGCGCTTGTCGCCGGA 792
Db 289 GTGGGCTCTGAGCTGCCCTTACGACCCAGCTTGTCGCCGCGCTTGTCGCCGCGCTTGTCGCCGGA 230
QY 793 GAGACCCCGCGCGCTTACGACCCAGCTTGTCGCCGCGCTTGTCGCCGCGCTTGTCGCCGCGGAT 852
Db 229 GAGACCCCGCGCGCTTACGACCCAGCTTGTCGCCGCGCTTGTCGCCGCGCTTGTCGCCGCGGAT 170
QY 853 TCCCGTGCACCTTGGGACCCAGAGTGGCCCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCCGCGG 912
Db 169 TCCCGTGCACCTTGGGACCCAGAGTGGCCCGCGCGCTCCCGCGCGCTCCCGCGCGCTCCCGCGG 110
QY 913 CCAGCAGCTCCAGAGCACTTACCCCGCGCGCTTGTCGCCGCGCTTGTCGCCGCGCTTGTCGCCGGA 972
Db 109 CCAGCAGCTCCAGAGCACTTACCCCGCGCGCTTGTCGCCGCGCTTGTCGCCGCGCTTGTCGCCGGA 50
QY 973 CCCTGGC 979
Db 49 CCCTGGC 43

RESULT 7
BM695925
LOCUS
DEFINITION
UI-E-CL1-afa-a-11-0-UI.r1 UI-E-CL1 Homo sapiens cDNA clone
UI-E-CL1-afa-a-11-0-UI 5', mRNA sequence.
BM695925
ACCESSION

```
VERSION BM695925.1 GI:19009183
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 599)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 202-272, >GC_rich#Low_complexity
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
source 1..599
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CL1-afa-a-11-0-UI"
/clone_lib="UI-E-CL1"
/tissue_type="human retina"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: p7T73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-CL1 is a normalized cDNA library containing the
following tissue(s): retina. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into p7T73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CCGCG. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
```

BASE COUNT 78 a 236 c 207 g 78 t

ORIGIN

Query Match 49.4%; Score 484; DB 14; Length 599;
Best Local Similarity 99.8%; Pred. No. 2.9e-211;
Matches 534; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 321 GGAGCGGCGCGCGCGGAGGAGGCTGAGATCAGCAGCGCGGCTCTGGCGCA 380
Db 1 GGAGCGGCGCGCGCGGAGGAGGCTGAGATCAGCAGCGCGGCTCTGGCGCA 60
QY 381 GCTGCTGGGCTCTGGGCGCGCGCGGAGGCTGAGATCAGCAGCGCGGCTCTGGCGCA 440
Db 61 GCTGCTGGGCTCTGGGCGCGCGCGGAGGCTGAGATCAGCAGCGCGGCTCTGGCGCA 120
QY 441 CCGCGAGCGGCTGAGCAGCGAGCTGCTGCGGCTCTGCTCCGCGCGCGGCTCTGACCTGCG 500
Db 121 CCGCGAGCGGCTGAGCAGCGAGCTGCTGCGGCTCTGCTCCGCGCGCGGCTCTGACCTGCG 180
QY 501 CGCCCTAGCAGCGGCTCTGCTCCGCGCGCGGCTCTGCTCCGCGCGCGGCTCTGACCTGCG 560
Db 1 CGCCCTAGCAGCGGCTCTGCTCCGCGCGCGGCTCTGCTCCGCGCGCGGCTCTGACCTGCG
```

```
Db 181 GCCTAGCAGCGGCTCTGCTCCGCGCGCGGCTCTGCTCCGCGCGGCTCTGACCTGCGGCTC 240
QY 561 CCGGCTCTAGCAGCGGCTCTGCTCCGCGCGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGG 620
Db 241 CCGGCTCTAGCAGCGGCTCTGCTCCGCGCGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAG 300
QY 621 CGAGCTGAGCAGCGGCTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 680
Db 301 CGAGCTGAGCAGCGGCTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
QY 681 CTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740
Db 361 CTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
QY 741 TGAGCTGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 800
Db 421 TGAGCTGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
QY 801 GCGCGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 855
Db 481 GCGCGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 535
```

RESULT 8

BM720023 537 bp mRNA linear EST 01-MAR-2002

LOCUS UI-E-EJ0-ahu-d-07-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone

DEFINITION UI-E-EJ0-ahu-d-07-0-UI 5', mRNA sequence.

ACCESSION BM720023

VERSION BM720023.1 GI:19038987

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 537)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL discovery

MEDLINE 97044477

COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA

sequence: 95-165, >GC_rich#Low_complexity

Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

source 1..537

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-EJ0-ahu-d-07-0-UI"

/clone_lib="UI-E-EJ0"

/tissue_type="fetal eyes, lens, eye anterior segment,

optic nerve, retina, Retina Foveal and Macular, RPE and

Choroid"

/dev_stage="fetal and adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: p7T73-Pac (Pharmacia) with a

modified polylinker; Site_1: EcoR I; Site_2: Not I;

UI-E-EJ0 is a substracted cDNA library constructed

according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAGA; lens, CGAATAGGA; eye anterior segment, AATCCGCAAT; optic nerve, CCATAAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 71 a 229 c 169 g 68 t

Query Match 49.3%; Score 483; DB 14; Length 537;
Best Local Similarity 99.8%; Pred. No. 8.2e-211;
Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 431 TGGACGAGGACCGCGGCTGCAGCGAGCTGCGTGGGCTGCTCGCGCGCGCC 490
Db 4 TGGACGAGACCGCGGCTGCAGCGAGCTGCGTGGGCTGCTCGCGCGCGCC 63
QY 491 TTGACCTGCGCGCTAGCAGCGAGCTGTCGCCGCGCGCTCCCGCGCGCTCC 550
Db 64 TTGACCTGCGCGCTGCGAGCGAGCTGTCGCCGCGCGCTCCCGCGCGCTCC 123
QY 551 GACCCCGCGCGCTAGCAGCGAGCGCGCGCGCGCGCTGCTGAGGAGCGAGCG 610
Db 124 GACCCCGCGCGCTAGCAGCGAGCGCGCGCGCGCGCTGCTGAGGAGCGAGCG 183
QY 611 ACGAGACCGCGAGTGGACCGCGAGCTGTTAGTACTTGTGGGACGAGTTCGCG 670
Db 184 ACGAGACCGCGAGTGGACCGCGAGCTGTTAGTACTTGTGGGACGAGTTCGCG 243
QY 671 GAACGCGGACTCCGAGGGGTGCGACCGCGCGCGCGCTCCCGCGCGAGCACG 730
Db 244 GAACGCGGACTCCGAGGGGTGCGACCGCGCGCGCGCTCCCGCGCGAGCACG 303
QY 731 ATGTGGGCTGTAGCTGCGCCCTGAGGGCGTGTGGGGCGCTGCTGCTGTGAACGCC 790
Db 304 ATGTGGGCTGTAGCTGCGCCCTGAGGGCGTGTGGGGCGCTGCTGCTGTGAACGCC 363
QY 791 TAGAGACCGCGCGCGCGAGTGGCTGCGACCGCGCTTTGGACACCTGAGCACTGCCG 850
Db 364 TAGAGACCGCGCGCGCGAGTGGCTGCGACCGCGCTTTGGACACCTGAGCACTGCCG 423
QY 851 GATCCCTGACCTGGGACCGAGAGTGGCCCGCGCATCCGCGCACGAGACTGCTCC 910
Db 424 GATCCCTGACCTGGGACCGAGAGTGGCCCGCGCATCCGCGCACGAGACTGCTCC 483
QY 911 CGCAGACGCTCCAGACAACTTACCCCGCGCGCGAGCGCTCTCAACCCGAGGA 964
Db 484 CGCAGACGCTCCAGACAACTTACCCCGCGCGCGAGCGCTCTCAACCCGAGGA 537

RESULT 9
BM504127/c
LOCUS
DEFINITION
h21e07.x1 Human insulinoma Homo sapiens cDNA 3', similar to
TF:Q9UHG2 Q9UHG2 PROSAS PRECURSOR. ; mRNA sequence.

ACCESSION
BM504127
VERSION
BM504127.1
KEYWORDS
EST.

SOURCE
GI:18666535

ORGANISM
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 593)

Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,

Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,

Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,

Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas

TITLE JOURNAL COMMENT

M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Powers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 427.

FEATURES source

1. 593
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human Insulinoma"
/tissue_type="Insulinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pBluescript SK-; Site:1:
xhoI; Site:2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

BASE COUNT 80 a 190 c 248 g 74 t
ORIGIN

Query Match 48.8%; Score 478; DB 13; Length 593;
Best Local Similarity 100.0%; Pred. No. 1.6e-208;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 476 TGCTCCGCGCGCGCTTGACCTGCGCGCTAGCAGCCAGCTTGTCCCGCGCGCTCC 535
Db 478 TGCTCCGCGCGCGCTTGACCTGCGCGCTAGCAGCCAGCTTGTCCCGCGCGCTCC 419
QY 536 CGCGCGCGCGCTCCGACCGCGCGCGCTTACGACGAGCGCGCGCGCGCGATG 595
Db 418 CGCGCGCGCGCTCCGACCGCGCGCGCTTACGACGAGCGCGCGCGCGCGATG 359
QY 596 CTGAGGAGCGCGCGAGACACCGCGCGCTGGACCGCGAGCTTGTGAGTACTTGTCTGG 655
Db 358 CTGAGGAGCGCGCGAGACACCGCGCGCTGGACCGCGAGCTTGTGAGTACTTGTCTGG 299
QY 656 GACGAGTCTTGGGGAGCGCGGACTTCCGAGGGGGTGGCAGCGCGCGCGCGCTCCGCC 715
Db 298 GACGAGTCTTGGGGAGCGCGGACTTCCGAGGGGGTGGCAGCGCGCGCGCGCTCCGCC 239
QY 716 GTCCCGCGCGAGCGAGTGGGCTCTGAGCTGCCCTTGGAGGGCGTCTGGGGCGCTGC 775
Db 238 GTCCCGCGCGAGCGAGTGGGCTCTGAGCTGCCCTTGGAGGGCGTCTGGGGCGCTGC 179
QY 776 TGGTGTGAACGCTAGAGACCGCGCGCGCTGACAGTGTCTGACGCGCGCTTGTTCGCGAC 835
Db 178 TGGTGTGAACGCTAGAGACCGCGCGCGCTGACAGTGTCTGACGCGCGCTTGTTCGCGAC 119
QY 836 CCTGAGCACTGCCCGGATCCCGTGCACCCCTGGGACCGAGAGTGGCCCGCGCATCCGCC 895
Db 118 CCTGAGCACTGCCCGGATCCCGTGCACCCCTGGGACCGAGAGTGGCCCGCGCATCCGCC 59
QY 896 ACCAGGACTGCTCCCGCGCGAGCGTCCAGAGCAACTTACCGCGCGCGCGCGCGCTTC 953
|||||

```

Db 58 ACCAGGACTGCTCCCGGCACAGCTCCAGACAACTTACCCCGGCACGACCCCTC 1
RESULT 10
Bi711080/c 601 bp mRNA linear EST 11-MAR-2002
LOCUS i490d01.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5023128
DEFINITION 3' similar to TR:Q9UH2 Q9UH2 PROSAA5 PRECURSOR. ;, mRNA sequence.
ACCESSION Bi711080
VERSION Bi711080
KEYWORDS EST.
SOURCE Bi711080.1 GI:15686775
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 601)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scarsee,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other ESTs: i490d01.y1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -400p from Glibco
High quality sequence stop: 435.
FEATURES
source
1. 601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5023128"
/tissue_type="Human insulinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:
XhoI; Site:2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
BASE COUNT 84 a 184 c 237 g 96 t
ORIGIN
Query Match 48.4%; Score 474; DB 13; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.le-206;
Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 506 TAGCAGCCAGCTTCTCCCGCGCCGTCGCCGCGCGCTCGACCCCGCCCGG 565
Db 513 TAGCAGCCAGCTTCTCCCGCGCCGTCGCCGCGCGCTCGACCCCGCCCGG 454
Qy 566 TCTACGACGAGCCGCCCGCGCCGCGATGCTGAGGAGCGAGCGGACGACACCCGACG 625
Db 453 TCTACGACGAGCCGCCCGCGCGCGGATGCTGAGGAGCGAGCGGACGACACCCGACG 394

```

```

Qy 626 TGGACCCCGAGCTGTGAGGTACTTGTGGGACGAGATTCTTGGGGAGAGCGGACTCG 685
Db 393 TGGACCCCGAGCTGTGAGGTACTTGTGGGACGAGATTCTTGGGGAGAGCGGACTCG 334
Qy 686 AGGGGTGTCAGCCCGCGCCGCTCCGCGTCGCCGCGCGACACCATGTGGGCTCTGAGC 745
Db 333 AGGGGTGTCAGCCCGCGCCGCTCCGCGTCGCCGCGCGACACCATGTGGGCTCTGAGC 274
Qy 746 TGCCCCCTGAGGGCTGTGTGGGGGCGCTGCTGCTGTGAAAGCGCTAGAGACCCCGGCGC 805
Db 273 TGCCCCCTGAGGGCTGTGTGGGGGCGCTGCTGCTGTGAAAGCGCTAGAGACCCCGGCGC 214
Qy 806 CCCAGTGCCTGACGCGCGCTTTCGCCACCTGAGACTGCCCGGATCCCGTGCACCT 865
Db 213 CCCAGTGCCTGACGCGCGCTTTCGCCACCTGAGCACTGCCCGGATCCCGTGCACCT 154
Qy 866 GGGACCCAGAGTGCCTCCCGGCATCCGCCACAGACTGCTCCCGGACAGCTCCAG 925
Db 153 GGGACCCAGAGTGCCTCCCGGCATCCGCCACAGACTGCTCCCGGACAGCTCCAG 94
Qy 926 AGCAACTTACCCCGCGCGCGCTCTCACCGAGGATCCCTACCCCGCTGGC 979
Db 93 AGCAACTTACCCCGCGCGCGCTCTCACCGAGGATCCCTACCCCGCTGGC 40
RESULT 11
BM052768/c 515 bp mRNA linear EST 12-MAR-2002
LOCUS ie59g07.x2 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION cDNA clone IMAGE:5672268 3' similar to TR:Q9UH2 Q9UH2 PROSAA5
PRECURSOR. ;, mRNA sequence.
ACCESSION BM052768
VERSION BM052768
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 515)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scarsee,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 421.
FEATURES
source
1. 515
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5672268"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site:1: Not 1;
Site:2: Sal 1; Starting library constructed using

```

SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

```
BASE COUNT      68 a  155 c  215 g   77 t
ORIGIN
Query Match      48.2%; Score 472; DB 13; Length 515;
Best Local Similarity 100.0%; Pred. No. 9e-206;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 508 GCAGCCAGCTGTCCCGCGCGCGTCCCGCGCGCGCTCCAGACCCCGCGCGTCCCGCGTCC 567
Db 497 GCAGCCAGCTGTCCCGCGCGCGTCCCGCGCGCGCTCCAGACCCCGCGCGTCCCGCGTCC 438
QY 568 TACGACGACGCGCGCGCGCGTCTGAGGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 627
Db 437 TACGACGACGCGCGCGCGCGTCTGAGGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 378
QY 628 GACCCGAGCTGTGAGGTACTTCTGGAGCGGATTTCTCGGGAAGCGGAGTCCGAG 687
Db 377 GACCCGAGCTGTGAGGTACTTCTGGAGCGGATTTCTCGGGAAGCGGAGTCCGAG 318
QY 688 GGGGTGCGACGCGCGCGCGCTCCCGCGCGCGTCCCGCGCGCGTCCCGCGCGCGTCCGAGCTG 747
Db 317 GGGGTGCGACGCGCGCGCGCTCCCGCGCGCGTCCCGCGCGCGTCCCGCGCGCGTCCGAGCTG 258
QY 748 CCCCTGAGGCGTGTGGGGCGCTGCTGCTGTGAAGCCCTAGAGACCCCGCGCGCG 807
Db 257 CCCCTGAGGCGTGTGGGGCGCTGCTGCTGTGAAGCCCTAGAGACCCCGCGCGCG 198
QY 808 CAGTGCTGACGCGCGCGCTTTTGGCCACCTGAGCACTGCGCGGATCCCGTGCACCCCTGG 867
Db 197 CAGTGCTGACGCGCGCGCTTTTGGCCACCTGAGCACTGCGCGGATCCCGTGCACCCCTGG 138
QY 868 GACCCAGAAGTCCCGCGCATCCCGCGCGAGGAGTCTCCCGCGCGAGGAGTCCAGAG 927
Db 137 GACCCAGAAGTCCCGCGCATCCCGCGCGAGGAGTCTCCCGCGCGAGGAGTCCAGAG 78
QY 928 CAACTTACCCCGGCGACGAGCCTCTCACCCGAGGATCCCTACCCCTGCG 979
Db 77 CAACTTACCCCGGCGACGAGCCTCTCACCCGAGGATCCCTACCCCTGCG 26
```

```
RESULT 12
BI789216/c
LOCUS
DEFINITION
  I53409.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
  cDNA clone IMAGE:5670688 3' similar to TR:Q9UH2 Q9UH2 PROSAA5
  PRECURSOR. ; mRNA sequence.
BI789216
BI789216.1 GI:15816941
EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 518)
AUTHORS
  Melton D., Brown J., Kenty G., Permutt, A., Lee, C., Kaestner, K.,
  Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
  Hillier, L., Marra, M., Pape, D., Wyllie, T., Martin, J., Blistain, A.,
  Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
  M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T.,
  Jackson, Y. and Bowers, Y.
  Endocrine Pancreas Consortium
TITLE
```

JOURNAL COMMENT

Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohpc.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 435.

FEATURES

source

```
1..518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5670688"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript plasmid library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."
```

BASE COUNT 68 a 155 c 215 g 80 t
ORIGIN

Query Match 48.2%; Score 472; DB 13; Length 518;
Best Local Similarity 100.0%; Pred. No. 9.1e-206;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 508 GCAGCCAGCTGTCCCGCGCGCGTCCCGCGCGCGTCCAGACCCCGCGCGTCCCGCGTCC 567
Db 500 GCAGCCAGCTGTCCCGCGCGCGTCCCGCGCGCGTCCAGACCCCGCGCGTCCCGCGTCC 441
QY 568 TACGACGACGCGCGCGCGTCTGAGGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 627
Db 440 TACGACGACGCGCGCGCGTCTGAGGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 381
QY 628 GACCCGAGCTGTGAGGTACTTCTGGAGCGGATTTCTCGGGAAGCGGAGTCCGAG 687
Db 380 GACCCGAGCTGTGAGGTACTTCTGGAGCGGATTTCTCGGGAAGCGGAGTCCGAG 321
QY 688 GGGGTGCGACGCGCGCGCTCCCGCGCGCGTCCCGCGCGCGTCCCGCGCGCGTCCGAGCTG 747
Db 320 GGGGTGCGACGCGCGCGCTCCCGCGCGCGTCCCGCGCGCGTCCCGCGCGCGTCCGAGCTG 261
QY 748 CCCCTGAGGCGTGTGGGGCGCTGCTGCTGTGAAGCCCTAGAGACCCCGCGCGCG 807
Db 260 CCCCTGAGGCGTGTGGGGCGCTGCTGCTGTGAAGCCCTAGAGACCCCGCGCGCG 201
QY 808 CAGTGCTGACGCGCGCGCTTTTGGCCACCTGAGCACTGCGCGGATCCCGTGCACCCCTGG 867
Db 200 CAGTGCTGACGCGCGCGCTTTTGGCCACCTGAGCACTGCGCGGATCCCGTGCACCCCTGG 141
QY 868 GACCCAGAAGTCCCGCGCATCCCGCGCGAGGAGTCTCCCGCGCGAGGAGTCCAGAG 927
```

```

Db 140 GACCCAGAGTCCCCGCCATCCCGCCACCAGGACTGCTCCCGCCACGACGTCACAG 81
QY 928 CAACCTTACCCCGCCAGCAGCCCTCTACCCGAGGATCCCTACCCCTGGC 979
|||||
Db 80 CAACCTTACCCCGCCAGCAGCCCTCTACCCGAGGATCCCTACCCCTGGC 29
|||||

RESULT 13
BF724971/c
LOCUS
DEFINITION
  541 bp mRNA linear EST 05-JAN-2001
  bx10f06.xl Human Iris cDNA (Un-normalized, unamplified): BX Homo
  sapiens cDNA clone bx10f06 3', mRNA sequence.
ACCESSION
  BF724971
VERSION
  BF724971.1 GI:12040882
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 541)
  Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
  NEIBANK: EST analysis and bioinformatics for ocular genomics
  Invest. Ophthalmol. Vis. Sci. 41, (2000) In press
  Contact: Wistow G
  Section on Molecular Structure and Function
  National Eye Institute
  6/331, NIH, Bethesda, MD 20892-2740, USA
  Tel: 301 402 3452
  Fax: 301 496 0078
  Email: graeme@helix.nih.gov
  Plate: 10 row: f column: 06
  Seq primer: -21M13 forward primer (ABI).
  Location/Qualifiers
    1..541
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="bx10f06"
      /clone_lib="Human Iris cDNA (Un-normalized, unamplified):
      BX"
      /tissue_type="Iris"
      /dev_stage="Adult"
      /lab_host="EMDH10B"
      /note="Organ: Eye; Vector: pCMVSPORT6; Post-mortem iris
      tissue was pooled from 10 individuals ranging in age from
      4-80 years and RNA was extracted. From this pooled sample
      an aliquot of 60ug of total RNA yielded 2.17ug of mRNA. A
      directionally cloned cDNA library in the pCMVSPORT6 vector
      was constructed at Life Technologies, essentially
      following the protocols of the Superscript Plasmid System
      full details of which are contained in the manufacturer's
      instruction manual (http://www.lifetech.com/). First
      strand synthesis was carried out using a Not I
      primer-adaptor [5'-pGACTAGTCTAGATCGGCGGCCCT(T)15-3',
      1. Not I/blunt end inserts were cloned into the Not I/EcoR
      V sites in the vector. EST analysis was performed on the
      unamplified library at the NIH Intramural Sequencing
      Center (NISC)."]
  BASE COUNT 69 a 159 c 222 g 91 t
  ORIGIN

Query Match
Best Local Similarity 48.2%; Score 472; DB 12; Length 541;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 GCAGCCAGCTTGTCGCCGCCCGCCCTCCCGCGCGGCTCCGACCCCGCGCCCGGTC 567
|||||
Db 512 GCAGCCAGCTTGTCGCCGCCCGCCCTCCCGCGCGGCTCCGACCCCGCGCCCGGTC 453
|||||
QY 568 TACGACGAGCGGCCCGCGCGCGGATGCTGAGGAGGAGCGGACGACACCCGACGTG 627
|||||
Db 452 TACGACGAGCGGCCCGCGCGGATGCTGAGGAGGAGCGGACGACACCCGACGTG 393
|||||
QY 628 GACCCGAGCTGTTGAGTACTTGCTGGGACGATCTTGGGGAGCGCGGACTCCGAG 687

```

```

Db 392 GACCCGAGCTGTTGAGTGACTTGTGGACGGAATCTTTGCGGAAGCGCGACTCCGAG 333
|||||
QY 688 GGGGTGGACAGCCCGCGCGCTCGCGCTGCGCGGACGAGTGTGGCTCTGAGCTG 747
|||||
Db 332 GGGGTGGACAGCCCGCGCGCTCGCGCTGCGCGGACGAGTGTGGCTCTGAGCTG 273
|||||
QY 748 CCCCCTGAGGGCGTGTGCTGGGGCGTGTGCTGTGTAACGCTAGAGACCCCGCGGCC 807
|||||
Db 272 CCCCCTGAGGGCGTGTGCTGGGGCGTGTGCTGTGTAACGCTAGAGACCCCGCGGCC 213
|||||
QY 808 CAGGTGCTGCACGCGCTCTTGGCACCTGAGCACTCCCGGGATCCGCTGACCCCTGG 867
|||||
Db 212 CAGGTGCTGCACGCGCTCTTGGCACCTGAGCACTCCCGGGATCCGCTGACCCCTGG 153
|||||
QY 868 GACCCAGAGTCCCGCGCATCCCGCCACAGGACTGCTCCCGCCACGACGTCACAG 927
|||||
Db 152 GACCCAGAGTCCCGCGCATCCCGCCACAGGACTGCTCCCGCCACGACGTCACAG 93
|||||
QY 928 CAACCTTACCCCGCCAGCAGCCCTCTACCCGAGGATCCCTACCCCTGGC 979
|||||
Db 92 CAACCTTACCCCGCCAGCAGCCCTCTACCCGAGGATCCCTACCCCTGGC 41
|||||

RESULT 14
BM142527/c
LOCUS
DEFINITION
  562 bp mRNA linear EST 12-MAR-2002
  if35d05.xl Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
  cDNA clone IMAGE:5678265 3' similar to TR:09UH2 Q9UH2 PROSAA5
  PRECURSOR. ; mRNA sequence.
ACCESSION
  BM142527
VERSION
  BM142527.1 GI:17152594
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 562)
  Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
  Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
  Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
  Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
  ,M., Gibbons,M., McCann,R., Cole,R., Tsagarelshvili,R., Williams,T.
  , Jackson,Y. and Bowers,Y.
  Endocrine Pancreas Consortium
  Unpublished (2000)
  Other ESTs: if35d05.yl
  Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
  Endocrine Pancreas Consortium
  Harvard University, Howard Hughes Medical Institute
  Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
  MA 02138
  Tel: 617-495-1812
  Fax: 617-495-8557
  Email: dmelton@lohp.harvard.edu
  Library was constructed by Dr. Douglas Melton DNA sequencing by:
  Washington University Genome Sequencing Center For information on
  obtaining a clone please contact: Juliana Brown
  (brown@fas.harvard.edu) This sequence now available from the IMAGE
  consortium, for clone orders contact: info@image.llnl.gov
  Seq primer: -400P from GIBCO
  High quality sequence stop: 426.
  Location/Qualifiers
    1..562
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:5678265"
      /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
      /sex="Both"
      /tissue_type="Islets of Langerhans"
      /dev_stage="Adult"
      /lab_host="DH10B"
      /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;

```


|||||
Db 140 GACCCAGAGTCCCCCGCCATCCCGCCACCCAGGACTGCTCCCCCGCCAGCACGTCCAGAG 81
QY 928 CAACTTACCCCGCCAGCCAGCCCTCTCACCCGAGGATCCCTACCCCTGGC 979
|||||
Db 80 CAACTTACCCCGCCAGCCAGCCCTCTCACCCGAGGATCCCTACCCCTGGC 29

Search completed: April 12, 2003, 20:46:19
Job time : 1608 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 20:46:29 ; Search time 152 Seconds
(without alignments)
5649.653 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcgacagcagcagc.....gagatccctaccctccctggc 979

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 593429 seqs, 438583890 residues

Word size : 15

Total number of hits satisfying chosen parameters: 3821

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- Published Applications NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	979	100.0	979	10	US-09-803-589-5
2	583	59.6	1050	9	US-09-984-271-58
3	226	23.1	301	10	US-09-880-107-1701
c	32	3.3	349	10	US-09-960-352-71
4	32	3.3	1027	10	US-09-803-589-13
5	25	2.6	342	10	US-09-983-965-3684
6	25	2.6	372	10	US-09-983-965-2663
7	20	2.0	372	10	US-09-854-133-589
8	19	1.9	349	9	US-09-833-790-259
9	19	1.9	349	10	US-09-917-800A-799
c	10	1.9	380	10	US-09-983-965-4193
c	11	1.9	424	10	US-09-833-790-30
12	19	1.9	464	10	US-09-880-107-2439
c	13	1.9	476	10	US-09-981-876-24
14	19	1.9	796	9	US-09-148-545-24
15	19	1.9	796	9	US-09-754-949-1
c	16	1.9	942	10	US-09-920-923-33
17	19	1.9	972	9	US-09-905-291A-126
18	19	1.9	1210	9	US-09-902-853-126
19	19	1.9	1210	9	US-09-902-853-126

20	19	1.9	1210	9	US-09-907-824-126	Sequence 126, App
21	19	1.9	1210	9	US-09-907-841-126	Sequence 126, App
22	19	1.9	1210	9	US-09-904-011-126	Sequence 126, App
23	19	1.9	1210	9	US-10-028-072-311	Sequence 311, App
24	19	1.9	1210	9	US-09-906-742-126	Sequence 126, App
25	19	1.9	1210	9	US-10-121-049-311	Sequence 311, App
26	19	1.9	1210	9	US-10-123-904-311	Sequence 311, App
27	19	1.9	1210	9	US-10-140-470-311	Sequence 311, App
28	19	1.9	1210	9	US-09-906-838-126	Sequence 126, App
29	19	1.9	1210	9	US-09-907-613-126	Sequence 126, App
30	19	1.9	1210	9	US-09-907-942-126	Sequence 126, App
31	19	1.9	1210	9	US-10-175-746-311	Sequence 311, App
32	19	1.9	1210	9	US-10-176-918-311	Sequence 311, App
33	19	1.9	1210	9	US-10-176-921-311	Sequence 311, App
34	19	1.9	1210	9	US-10-137-865-311	Sequence 311, App
35	19	1.9	1210	9	US-10-140-474-311	Sequence 311, App
36	19	1.9	1210	9	US-09-904-820-126	Sequence 126, App
37	19	1.9	1210	9	US-09-904-859-126	Sequence 126, App
38	19	1.9	1210	9	US-09-909-204-126	Sequence 126, App
39	19	1.9	1210	9	US-10-142-431-311	Sequence 311, App
40	19	1.9	1210	9	US-10-143-114-311	Sequence 311, App
41	19	1.9	1210	9	US-09-904-786-126	Sequence 126, App
42	19	1.9	1210	9	US-09-906-646-126	Sequence 126, App
43	19	1.9	1210	9	US-09-906-700-126	Sequence 126, App
44	19	1.9	1210	9	US-10-140-002-311	Sequence 311, App
45	19	1.9	1210	9	US-09-902-903-126	Sequence 126, App

ALIGNMENTS

RESULT 1
US-09-803-589-5
; Sequence 5, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 979
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(837)
US-09-803-589-5

Query Match 100.0%; Score 979; DB 10; Length 979;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 979; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAATTCCGGCAGGAGCCAGCCAGTCCGCCSGYMCRRGCCCGGCTCGCTGGGGCAGCATG	60
Db	1	GAATTCCGGCAGGAGCCAGCCAGTCCGCCSGYMCRRGCCCGGCTCGCTGGGGCAGCATG	60
QY	61	GGGGGTCCGCGCTGCTCTTGGGGCCGCGGGCCGGGGCGTCCGCCCTTTTGGTGTGCTG	120
Db	61	GGGGGTCCGCGCTGCTCTTGGGGCCGCGGGCCGGGGCGTCCGCCCTTTTGGTGTGCTG	120
QY	121	CTGCTCGGGCTGTTTGGCCGCCGCCGCCGGCTCTGCGCGGGCCGGTAAAGAGACCCCG	180
Db	121	CTGCTCGGGCTGTTTGGCCGCCGCCGCCGGCTCTGCGCGGGCCGGTAAAGAGACCCCG	180
QY	181	GGCTAAGCGCAGCTCTCCGCCCTTGGCTAGACATGGGCGTCTCTCGCGCTTCCGGCGG	240
Db	181	GGCTAAGCGCAGCTCTCCGCCCTTGGCTAGACATGGGCGTCTCTCGCGCTTCCGGCGG	240
QY	241	TCAGTGCCTCCGAGGTGAGCGGGGGGGGTGTCAGAGAGTGGCGGGCGTGGCGCAT	300
Db	241	TCAGTGCCTCCGAGGTGAGCGGGGGGGGTGTCAGAGAGTGGCGGGCGTGGCGCAT	300
QY	301	CTGCTGGAGCGCAGCTCAGAGCGGGCGCGGGCCAGCGCAGAGAGCTGAGGATCAG	360
Db	301	CTGCTGGAGCGCAGCTCAGAGCGGGCGCGGGCCAGCGCAGAGAGCTGAGGATCAG	360
QY	361	CAGCGCGCTCTCGCGCAGCTGCTCGCGTCTGGGGCGCCGCCCGCAACTCTGATCG	420
Db	361	CAGCGCGCTCTCGCGCAGCTGCTCGCGTCTGGGGCGCCGCCCGCAACTCTGATCG	420
QY	421	GCTCTGGGCTTGGAGCAGCACCCGACGGCCCTGACGAGAGAGTCTCGCTCGCGCTCTGCTC	480
Db	421	GCTCTGGGCTTGGAGCAGCACCCGACGGCCCTGACGAGAGAGTCTCGCTCGCGCTCTGCTC	480
QY	481	CGCGCCGCTTACCTCGCCGCTAGCAGCCACGCTTGCTCCCGCGCCGCTCCCGCC	540
Db	481	CGCGCCGCTTACCTCGCCGCTAGCAGCCACGCTTGCTCCCGCGCCGCTCCCGCC	540
QY	541	CGCGCGTCCGACCCCGGCCCGCGTCTACGACAGCGGCCGCGGGCCGGATGCTGAG	600
Db	541	CGCGCGTCCGACCCCGGCCCGCGTCTACGACAGCGGCCGCGGGCCGGATGCTGAG	600
QY	601	GAGCAGCGCAGCAGACACCCGACGTGGACCCGAGCTGTGAGGTACTTGTGGGACGG	660
Db	601	GAGCAGCGCAGCAGACACCCGACGTGGACCCGAGCTGTGAGGTACTTGTGGGACGG	660
QY	661	ATTCTTGGGGAAGCGGGACTCCGAGGGGTGGCAGCCCGCGCGCTCCCGCGTGC	720
Db	661	ATTCTTGGGGAAGCGGGACTCCGAGGGGTGGCAGCCCGCGCGCTCCCGCGTGC	720
QY	721	GCGACACAGTCTGGGCTGTAGCTGCCCTGAGGGGTGCTGGGGGCGCTGCTGCGT	780
Db	721	GCGACACAGTCTGGGCTGTAGCTGCCCTGAGGGGTGCTGGGGGCGCTGCTGCGT	780
QY	781	GTAAACGCCTAGAGACCCCGCGCCACAGGTGCTTGCACGCGGCTCTTGGCACCCCTGA	840
Db	781	GTAAACGCCTAGAGACCCCGCGCCACAGGTGCTTGCACGCGGCTCTTGGCACCCCTGA	840
QY	841	GCACTGCCGGATCCCGTGCACCCCTGGGACCCAGAGTGCCTCGCGGCATCCCGCCAC	900
Db	841	GCACTGCCGGATCCCGTGCACCCCTGGGACCCAGAGTGCCTCGCGGCATCCCGCCAC	900
QY	901	GACTGCTCCCGCAGCAGTCCAGAGCAACTTACCCCGCGCAGCAGCGCTCTCACCG	960
Db	901	GACTGCTCCCGCAGCAGTCCAGAGCAACTTACCCCGCGCAGCAGCGCTCTCACCG	960
QY	961	AGGATCCCTACCCCTGGC	979
Db	961	AGGATCCCTACCCCTGGC	979

RESULT 2

RESOL 2
US-09-984-271-58

03 03 304 271 30
; Sequence 58, Application US/09984271

```

: Publication No. US20030040088A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Rosen et al.
:
: TITLE OF INVENTION: 71 Human Secreted Proteins
:
: FILE REFERENCE: pz030p1
:
: CURRENT APPLICATION NUMBER: US/09/984, 271
:
: CURRENT FILING DATE: 2001-10-29
:
: PRIOR APPLICATION NUMBER: 09/482,273
:
: PRIOR FILING DATE: 2000-01-13
:
: PRIOR APPLICATION NUMBER: PCT/US99/15849
:
: PRIOR FILING DATE: 1999-07-14
:
: PRIOR APPLICATION NUMBER: 60/092,921
:
: PRIOR FILING DATE: 1998-07-15
:
: PRIOR APPLICATION NUMBER: 60/092,922
:
: PRIOR FILING DATE: 1998-07-15
:
: PRIOR APPLICATION NUMBER: 60/092,956
:
: PRIOR FILING DATE: 1998-07-15
:
: NUMBER OF SEQ ID NOS: 267
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 58
:
: LENGTH: 1050
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: US-09-984-271-58

```

Query Match

	QY	346	GAGGCTGAGATCAGACGGCGCGGTCTTGGCCAGAGTGCTGCCTCTTGGGGCGCCCC	405
	Db	339	GAGGCTGAGATCAGACGGCGCGGTCTTGGCCAGAGTGCTGCCTCTTGGGGCGCCCC	398
	QY	406	CGCAACTCTCATCCGGCTCTGGGCTTTGACGACGACCCGACGCGCTGCAGGCGAGTC	465
	Db	399	CGCAACTCTCATCCGGCTCTTGGGCTTTGACGACGACCCGACGCGCTGCAGGCGAGTC	458
	QY	466	GCTCGGCTCTGTCTCGCGCCCBCCTTTGACCTTGGCGCCTTAGCACGCCAGCTGTGCC	525
	Db	459	GCTCGGCTCTGTCTCGCGCCCBCCTTAGCTTGGCGCCTTAGCACGCCAGCTGTGCC	518
	QY	526	GGCCCGTCCCGCGCGGGGCTCCGACCOCGCGCCCGCTCTACGACGACGGCCCGCG	585
	Db	519	GGCCCGTCCCGCGCGGGGCTCCGACCOCGCGCCCGCTCTACGACGACGGCCCGCG	578
	QY	586	GGCCCGGATGCTGAGGAGCAGCGACGACACCCGACBTGGACCCCGAGCTGTGTAGG	645
	Db	579	GGCCCGGATGCTGAGGAGCAGCGACGACACCCGACBTGGACCCCGAGCTGTGTAGG	638
	QY	646	TACTTCTGGGACGGATTCTTGGGGNAAGCGCGGACTCCGAGGGGTGGCAGCCCGCGC	705
	Db	639	TACTTCTGGGACGGATTCTTGGGAAGCGCGGACTCCGAGGGGTGGCAGCCCGCGC	698
	QY	706	CGCTCTCGCGCTCGCGCCAGCACGATGTGGGCTCTGAGTGCCTCTGAGGCGGTGCTG	765
	Db	699	CGCTCTCGCGCTCGCGCCAGCACGATGTGGGCTCTGAGTGCCTCTGAGGCGGTGCTG	758
	QY	766	GGGGCGCTGCTGCGTGTGAACGCCCTAGAGACCCCGCGCCAGTGCCTGCACGCCGC	825
	Db	759	GGGGCGCTGCTGCGTGTGAACGCCCTAGAGACCCCGCGCCAGTGCCTGCACGCCGC	818
	QY	826	CTCTTGGCCACCTGAGCACTGCCCGGATCCCGTGGACCTTGGGACCCAGAAGTGC	885
	Db	819	CTCTTGGCACCTTGGCACTTGCCCGGATCCCGTGGACCTTGGGACCCAGAAGTGC	878
	QY	886	CCATCCCGCCACCAAGACTGCTCCCGCCAGACAGTCCAGAGCAACTTACCCGGCGCAG	945
	Db	879	CCATCCCGCCACCAAGACTGCTCCCGCCAGACAGTCCAGAGCAACTTACCCGGCGCAG	938
	QY	946	CAGCCCTCTCACCCGAGGATCCCTACCCCTGGC	979
	Db	939	CAGCCCTCTCACCCGAGGATCCCTACCCCTGGC	972

```

US-09-960-352-71
; Sequence 71, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Wyatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 71
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB3057-002-01-K1-A9

```

; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 3684
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 43-LIB3058-043-Q1-K1-C12
US-09-983-965-3684

Query Match 2.6%; Score 25; DB 10; Length 342;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 GAGGATCAGCAGCGCGCTCTGG 376
|||||
Db 4 GAGGATCAGCAGCGCGCTCTGG 28

RESULT 7

US-09-983-965-2663
; Sequence 2663, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Machalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2663
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 10-LIB3058-043-Q1-K1-C5
US-09-983-965-2663

Query Match 2.0%; Score 20; DB 10; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 GCAGCCAGCTGTCCCGC 527
|||||
Db 127 GCAGCCAGCTGTCCCGC 146

RESULT 8

US-09-854-133-589
; Sequence 589, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11

; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 589
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-133-589

Query Match 1.9%; Score 19; DB 9; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCAGGAGCCAG 19
|||||
Db 1 GAATTCGGCAGGAGCCAG 19

RESULT 9

US-09-833-790-259
; Sequence 259, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 259
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-790-259

Query Match 1.9%; Score 19; DB 10; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGCAGGAGCCAG 19
|||||
Db 1 GAATTCGGCAGGAGCCAG 19

RESULT 10

US-09-917-800A-799/c
; Sequence 799, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Etashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15

```
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 799
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1136514
US-09-917-800A-799

Query Match      1.9%; Score 19; DB 10; Length 380;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 AATTCGGCAGGCGCCAGC 20
Db      380 AATTCGGCAGGCGCCAGC 362

RESULT 11
US-09-983-965-4193/c
; Sequence 4193, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4193
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 59-LIB3058-051-Q1-K1-G4
US-09-983-965-4193

Query Match      1.9%; Score 19; DB 10; Length 424;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      742 GAGCTGCCCTGAGGGCG 760
Db      108 GAGCTGCCCTGAGGGCG 90

RESULT 12
US-09-833-790-30
; Sequence 30, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tonglong
; APPLICANT: Secrist, Heather
```

```
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapien
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 N20198
US-09-833-790-30

Query Match      1.9%; Score 19; DB 10; Length 464;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAATTCGGCAGGCGCCAG 19
Db      1 GAATTCGGCAGGCGCCAG 19

RESULT 13
US-09-880-107-2439/c
; Sequence 2439, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 2439
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 N20198
; NAME/KEY: unsure
; LOCATION: (1)..(476)
; OTHER INFORMATION: n = a o c o r g o r t
US-09-880-107-2439

Query Match      1.9%; Score 19; DB 10; Length 476;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      107 TTTTGTGCTGCTGCTGCT 125
Db      445 TTTTGTGCTGCTGCTGCT 427

RESULT 14
US-09-981-876-24
; Sequence 24, Application US/09981876
; Patent No. US20020164669A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/981,876
```



```

; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,590
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,594
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,589
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,593
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,614
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,578
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/047,501
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,670
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/056,632
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,876
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,881
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,909
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,875
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,862
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,887
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,908
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/048,964
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/057,650
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/056,884
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 796

Query Match      1.9%; Score 19; DB 9; Length 796;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 TGCTGCTGCTGCTCGGCCT 131
      |||||
Db 62 TGCTGCTGCTGCTCGGCCT 80

RESULT 15
US-09-148-545-24
; Sequence 24, Application US/09148545
; Publication No. US20030027132A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/148,545
; CURRENT FILING DATE: 1998-09-04
; EARLIER APPLICATION NUMBER: PCT/US98/04482
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,161
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
```

; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 05-Sep-1997
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 796

Query Match 1.9%; Score 19; DB 9; Length 796;
Best Local Similarity 100.0%; Pred. No. 4.3;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 TCGTCTGCTGCTGGCCT 131

Db 62 TCGTCTGCTGCTGGCCT 80

Search completed: April 12, 2003, 21:41:32
Job time : 162 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	146	11.1	913	2	B97369	translation initia	
2	146	11.1	913	2	A82587	translation initia	
3	134	10.2	553	2	C75318	hypothetical prote	
4	130.5	9.9	589	2	G87485	hypothetical prote	
5	127.5	9.7	460	2	T33110	hypothetical prote	
6	124.5	9.4	815	2	B36708	extracellular sign	
7	123	9.3	801	2	T29018	hypothetical prote	
8	123	9.3	1238	2	T03465	probable exonuclea	
9	121.5	9.2	590	1	T35297	probable dihydrol	
10	119.5	9.1	422	2	S23357	glial growth facto	
11	119.5	9.1	1460	1	EDBEIF	immediate-early pr	
12	118.5	9.0	1298	1	EDBE75	immediate-early pr	
13	118	8.9	1446	1	A45344	immediate-early pr	
14	117	8.9	643	1	F0LJLK	gag polyprotein -	
15	116.5	8.8	1677	2	T14267	Xin protein, stage	
16	116	8.8	548	2	S59133	E2S2 repressor fac	
17	115.5	8.8	998	2	T35745	probable ATP-bind	
18	115	8.7	2240	2	T37057	probable multi-dom	
19	114	8.6	976	2	G87389	Tona-dependent rec	
20	113.5	8.6	559	2	F83283	precorrin-3 methyl	
21	113.5	8.6	1089	2	T36663	protein kinase, tr	
22	113.5	8.6	3164	1	WMBEH6	U36 protein - hum	
23	112.5	8.5	413	2	S61305	cych protein - Par	
24	112.5	8.5	933	1	QRHUP	progesterone recep	
25	112	8.5	817	2	T03852	protein phosphatas	
26	112	8.5	1147	2	D87295	smc protein [impor	
27	111	8.4	846	2	S2418	grp-binding regula	
28	111	8.4	1607	2	G87259	conserved hypothet	
29	111	8.4	1870	2	S37671	MHC class III hist	

translation initiation factor IF-2 [imported] - Agrobacterium tumefaciens (s
C; Species: Agrobacterium tumefaciens
C; Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gilliet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AB2587
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <KUR>
A:Cross-references: GB:AE006588; PIDN:AA141112.1; PID:g17738404; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: infB
A:Map position: circular chromosome
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu hom

Query Match 11.1%; Score 146; DB 2; Length 913;
Best Local Similarity 25.2%; Pred. No. 0.034;
Matches 65; Conservative 34; Mismatches 79; Indels 80; Gaps 12;

QY 29 PPALCARP-----VKEPGL-----SAASPPLAETGAPRRFRSVPGE-----AAGAV 72
Db PAPVAARPOAPRIHQPGQOORPGSSQSRGSAPOQADPRGNVLDLSAGEM 131
QY 73 QELARALAHLEA-----ERGERARAEQAEQARVLAQLLRVWGAPRNSDPALGLD 126
Db 132 EARRRA---LMEAQADVVEAKQRAEDARRKVEEQRITAE---KMEANRAAEAEAAAK 186
QY 127 DDPDPAQAQLARALLRDLPAALAAQLVPAVPVPAALRPVVDGPGAGDAEAGDE 186
Db 187 VAASOPAAEV-----RAEPASERPAAPAPRTDARPO-----SAAAPASRAPA--- 231
QY 187 TPDVDPPELLRYLLGRILAGSADSEGAAPRLRRADHD-----VGSFLPPEGVLGAL 239
Db 232 TFD-----AAPGRRTGGDEDDRGAVRGRGSSLPARG----- 264
QY 240 LRVKRLTAPQVPAARRL 257
Db 265 ----KVVAAPAKPAARL 278

RESULT 3
C75318
hypothetical protein - *Deinococcus radiodurans* (strain R1)
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75318
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; PMID:20036996; PMID:10567266
A:Accession: C75318
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-553 <WHI>
A:Cross-references: GB:AE002044; GB:AE000513; NID:g6459872; PIDN:AAF11639.1; PID:g645986
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2090
A:Map position: 1

Query Match 10.2%; Score 134; DB 2; Length 553;
Best Local Similarity 25.0%; Pred. No. 0.12;
Matches 70; Conservative 24; Mismatches 108; Indels 78; Gaps 11;

QY 24 GLFRPPPALCARPVEKPRG-----LSAASPP-----LAET 53
Db 32 GTLAEAPVQIPIVRRPAGQVPTTSIALSGVEPPAPRAEVQVRAVRDPSRRTPTAAEP 91

QY 54 GAPRRFRSVPGEAAGAVQELARALAHLEAERQERAR-----AFAQAEADQOAR 104
Db VAPTRTSRGARPTPAETVPSAPVQAPVAEAREPKPRPTTAAAPVPAASPAETEAA 151
QY 105 VLAQLLRVWGAPRNSDPALGLDD---DPDPAQAQLARALLRDLPAALAAQ---LVPAP 158
Db 152 PAS-----APRSGWLSSALNDWLKEPEAPAPAPREAPARQVQERSTATQTVQAAAP 204
QY 159 VPAALRPVVDGPGAGDAEE---AGDETDPDPELLRYLLGRILAGSADSEGAAP 215
Db 205 KPAPA---PQP---DSDPSLPTLQELASDLRLDLPVELLERLWEQEQAAQOEPEAPAP 259
QY 216 RLRLRAADHDVSGELPPEGVLGALLRVKRLTAPQVPAAR 255
Db 260 ----RAPVRVAAALSP-----APSPAAPR 281

RESULT 4
G87485
hypothetical protein CC1908 [imported] - *Caulobacter crescentus*
C:Species: *Caulobacter crescentus*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: G87485
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: AB7249; PMID:21173698; PMID:11259647
A:Accession: G87485
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-589 <STO>
A:Cross-references: GB:AE005673; NID:g13423359; PIDN:AAK23883.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1908

Query Match 9.9%; Score 130.5; DB 2; Length 589;
Best Local Similarity 29.8%; Pred. No. 0.22;
Matches 82; Conservative 15; Mismatches 91; Indels 87; Gaps 15;

QY 45 AASPPLAETGAPRRFRSVPGEAAGAVQELARAL-AHLEAERQERARA----- 93
Db 182 AAAPPSAVGAAP-----PMVAEGDLKAALWLRLSALSALAKAEVSSGAVAPHVGSVADGG 237
QY 94 ----EAQAEQOARVLAQLLRVWGAPRNSDPALG---LDDDPDAPAA---QLARALLRA 143
Db 238 GADGPDGPGMTHQARANGSSSA-GMASSPPAQSRSLAADAPDAPANTSSPTARAL--- 293
QY 144 RLDPALAAQLVPAVPVPAALRPVVDGPGAGDAEAGDETDPDPEL-LRY----- 197
Db 294 -----PAGAPVDAEDPPPARVCAPPSSG-----SPTHADADELAVRFGAFVA 335
QY 198 -----LLGRILAGSA-----DSEG-----VAAPRRLLR-----AADHDVGS 228
Db 336 PPQKTAERSPGKVALGASLPALVOLPRDGEPPATEPAEPRLMTRGYGAPVEDVRS 395
QY 229 ELPPGVLGALLRVKRLT-TP-----APQVPAARLL 258
Db 396 KTPPPPYAGPTAGQRPEPTPSLSWAPEAMARRLL 430

RESULT 5
T33110
hypothetical protein C18H7.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T33110
R:Tin-Wollam, A.; Fronick, W.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of *C. elegans* cosmid C18H7.
A:Reference number: Z21284

A:Accession: T33110
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-460 <TIN>
A:Cross-references: EMBL:AF067607; PIDN:AA017641.1; GSPDB:GN00022; CESP:C18H7.3
A:Experimental source: strain Bristol N2; clone C18H7
C:Genetics:
A:Gene: CESP:C18H7.3
A:Map position: 4
A:Introns: 84/1
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 9.7%; Score 127.5; DB 2; Length 460;
Best Local Similarity 30.0%; Pred. No. 0.27;
Matches 64; Conservative 10; Mismatches 84; Indels 55; Gaps 8;
QY 29 PP-----ALCARPVKEPRGLSASPLAETGAPRRFRSVPGRCEAGAGVOELARALAHLEA 85
Db 275 PPSNALGAGGGAEPAG-----AAPEAAAAAPEAAPEAAAGAGGAGGAEAPAGAAAPDAAA 330
QY 86 ERQERARAEQAEDQOQARVLAQLLRVWAGPRNSDPALGLDD-----DPDPAQAQLARALL 141
Db 331 AAPEAAPAEAPAAE-----GAGGGAEPAGAAPDAAAAAPEAPAEAPAAE 377
QY 142 RA--RLDPAALAAQLVPAPVPAALRPVDDGPDAGPDAEEAGDETDPDPELLRYLL 199
Db 378 GAGGGAEPAGAAAPDAAAPEAAPE-----AAPAEAGAG----- 414
QY 200 GRILAGSADSEGVAAAPRLRRADHDVGSLEPP 232
Db 415 -----GGAEPAG-AAPEAAAAAPGAGGGEAPP 441

RESULT 6
B56708
extracellular signal-regulated kinase 5 - human
C:Species: Homo sapiens (man)
C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 24-Sep-1999
C:Accession: B56708
R:Zhou, G.; Bao, Z.Q.; Dixon, J.E.
J. Biol. Chem. 270, 12665-12669, 1995
A:Title: Components of a new human protein kinase signal transduction pathway.
A:Reference number: A56708; MUID:95279403; PMID:7759517
A:Accession: B56708
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-815 <ZHO>
A:Cross-references: GB:U25278; NID:g837260; PIDN:AAA81381.1; PID:g837261
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
F:52-315/Domain: protein kinase homology <KIN>
F:60-68/Region: protein kinase ATP-binding motif

Query Match 9.4%; Score 124.5; DB 2; Length 815;
Best Local Similarity 22.7%; Pred. No. 0.76;
Matches 67; Conservative 31; Mismatches 116; Indels 81; Gaps 11;
QY 5 PLWMPRAGGVGLLVLLGLFRPPA--LCARPVKEPRGLSASPLAETGAPRRFRS 62
Db 419 PSPWAPSGD-----CAMESPPAPPCCGAPDITDLTQPPPPVSEPPAPPKKDXA 469
QY 63 VPRGEAAGAGVQELARALAHLL-----EAERQERARAEQAED 100
Db 470 ISDNWTKAALKAAKLSRSLRDGSPAPLEAPRKPVTAQEREREERRRRQERAKE 529
QY 101 QOARVLAQLLRVWAGPR-----NSDPALGL-----DDPDPAQAQLARALLRDLPAALAAQ 153
Db 530 REXRQERERKERAGAGSGPSTDLGLVLSND-----RSLERWTMRAR--PAAPALT 583
QY 154 LVPAPVPAALRRPPVDDGPDAGPDAEEAGDETDPDPELLRYLLGRILAGSADSEGA 213
Db 584 SVAPAPAPPTPTTTPVQPTSPPPGFLAQPTGPQ-----PQSAGSTSGVPVQPCPPPG-- 636

QY 214 APRRLRAADHDVGSLEPPGEGVIGALLRVKRLKLETPAQ-----VPARRLLPP 260
Db 637 -----PAPHTGPPGB-----IPVAPPQIATSTSLAAQSLVPP 671
RESULT 7
T29018
hypothetical protein ZK84.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T29018
R:Kirsten, J.
submitted to the EMBL Data Library, April 1995
A:Description: The sequence of C. elegans cosmid ZK84.
A:Reference number: Z20553
A:Accession: T29018
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-801 <KIR>
A:Cross-references: EMBL:U23181; PIDN:AA048204.1; GSPDB:GN00020; CESP:ZK84.1
A:Experimental source: strain Bristol N2; clone ZK84
C:Genetics:
A:Gene: CESP:ZK84.1
A:Map position: 2
A:Introns: 22/2; 45/3; 108/1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match 9.3%; Score 123; DB 2; Length 801;
Best Local Similarity 28.5%; Pred. No. 0.93;
Matches 73; Conservative 14; Mismatches 107; Indels 62; Gaps 12;
QY 28 PPPALCARPVKEPRGLSASPLAETGAPRRFRSVPGRGAA-----GAVQELARALAH 81
Db 507 PAPAPIEAPATDAATLETAPAAAE-----PAPAAEAAGYDAPSSVPEETPAPAP 557
QY 82 LLE-----AERQERARAEQAEDQOQARVLAQLLRVWAGPRNSD---PALGLDDDP-DAP 132
Db 558 ADETPAPAAAEETPAPAAAEETPAPAAAEETPAPAAAEETPAPAAAEETPAPAAAEETPAPAP 617
QY 133 AAQLARALLRA-RLDPA--ALAAQLVPAPVPAALRP--P-----PVYDDGPA-GPDA 180
Db 618 AAETPAPAAVEETPAPAAVEETPAPAAVEETPAPAAVEETPAPAAVEETPAPAAVEETPAPAA 677
QY 181 EE-----AGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPP 234
Db 678 EETPAPAAAEETPAPAP-----AAETPAPAAAEETPAPAAAEETPAPAAAEAPAPA 724
QY 235 VLGALLRVKRLKLETPAP 250
Db 725 PAAP-----ETPAP 733

RESULT 8
T03465
Probable exonuclease (EC 3.1.15.-) - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Jun-2000
C:Accession: T03465
R:Vicek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SBI
A:Reference number: 214955; MUID:9740404; PMID:9256491
A:Accession: T03465
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1238 <VLC>
A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AA016118.1; PID:g3128266
C:Genetics:
A:Gene: sbcc
A:Map position: 1
C:Superfamily: chromosome segregation protein SMC1
C:Keywords: DNA repair; exonuclease; hydrolase

```

QY      249  APOVPARRLLP 259
      || || : |
Db      335  APAAPAAKKAP 345

RESULT 10
S32357
glial growth factor - human
C:Species: Homo sapiens (man)
C:Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
R:Accession: S32357
R:Marchionni, M.A.; Goodearl, A.D.J.; Chen, M.S.; Bermingham-McDonogh, O.; Kirk, C.;
  les, I.; Davis, J.B.; Hsuan, J.J.; Totty, N.F.; Otsu, M.; McBurney, R.N.; Waterfield,
  Nature 362, 312-318, 1993
A:Title: Glial growth factors are alternatively spliced erbB2 ligands expressed in th
A:Reference number: S32357; MUID:93205115; PMID:8096067
A:Accession: S32357
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-422 <MAR>
A:Cross-references: GB:112260; NID:g292047; PIDN:RAB59622.1; PID:g292048
C:Superfamily: human heregulin; EGF homology
F:363-402/Domain: EGF homology <EGF>

Query Match      9.1%; Score 119.5; DB 2; Length 422;
Best Local Similarity 28.0%; Pred No. 0.02.

```

```

QY 2 AGSPLLMGPRAGGVG-----LVLVLLGLFRPPALCARPVKEPRGLSA--ASP 48
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 SGRD---GPRQRPGSAARSSPPLPLPLLLGLLTAALPGAAAGNEAAPAGASVCVYSSP 66
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 49 PLAETGAPRRFRSSVPRGEAAGAVQELARALAHLLAEERQERARAEAEEDQOARVLAQ 108
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 P-----SVGSVQELAAQAAVVIEGKVPQRR--OOGALDRKAAAAAG 106
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 109 LLRWGAPRNSDPALGLDDDPDAPAAQLARALLRARLDPAALAAQLPVPVPAALPRP 168
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 107 EAGAWGG-----DREPPAA-----GPRALGP--PAEEPLLAANGTV 140
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 169 PVYDDGDPAGPDAAEAGDETP 188
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 141 PSWTPAPV-PSAGEPGGEAP 159
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 11
EDBEIF
immediate-early protein IE180 - suid herpesvirus 1 (strain Indiana-Funkhauser)
C;Species: suid herpesvirus 1
C;date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 16-Feb-1997
C;Accession: S04713
R;Cheung, A.K.

```

Nucleic Acids Res. 17, 4637-4646, 1989

A;Title: DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies A virus
A;Reference number: S04713; MID:89315207; PMID:2546124
A;Accession: S04713
A:Molecule type: DNA
A:Residues: 1-1460 <CH>
C:Superfamily: herpesvirus immediate-early protein IEL75
C:Keywords: DNA binding; early protein; transcription regulation

Query Match 9.1%; Score 119.5; DB 1; Length 1460;
Best Local Similarity 25.2%; Pred. NO. 2.9;
Matches 75; Conservative 24; Mismatches 110; Indels 89; Gaps 14;

Oy: 3 GSPLLKGP--RAGVGLLVLLLGLPRPPALCARPVKEPRLGSLSAASPLLAETCAPRRFR 60
db: 331 GAPLQRPPIRRRRAGEAL-----RRRGFSSSSGGSDDLSPARS 371

Oy: 61 RSVPRGEACAVQELARALAHLLEAEQRERARAQAQA-EQQOARVLQAQLLRVYWGAPRNS 119
db: 372 PSAPRAPAAAAA---ARRRSASSSSSSSSSSSSSSSSSSSSSBEDEGVRCGAPTARAGPPP--- 426

QY 120 DPALGLDDDPAPAAQLARALLRARLDPAALAAQLVPAPVPAALRRPP----- 169
 Db 427 -----SPPAPAAAPRPSASSASATSSAAASPAPAPEFA-----RPPRRKRSTNNH 473
 QY 170 --VYDDGPA---GDAEAGDETPDVDPQL---LRYLLGRILAGSADS-BGV-----A 213
 Db 474 LSLMADGPPPTDGLPLTGLPEWPGSPDPADGRVY-----GGAGDSREGLWDEDDVRO 527
 QY 214 APRRLRAA-----DHDVGSLEPPEGVGLGALLRVKRLTETAPQVPA 256
 Db 528 AARYRAAGPVPVFIPEMGDSRKQHEALVRLIYSGAAGEAM--SWLQNPQMOPDQR 583

RESULT 12

EDBE75
 Immediate-early protein IE175 - human herpesvirus 1
 C:Species: human herpesvirus 1
 C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jun-2000
 C:Accession: A23510
 R:McGeoch, D.J.; Dolan, A.; Donald, S.; Brauer, D.H.K.
 Nucleic Acids Res. 14, 1727-1745, 1986
 A:Title: Complete DNA sequence of the short repeat region in the genome of herpes simplex
 A:Reference number: A23510; MUID:86148504; PMID:3005980
 A:Accession: A23510
 A:Molecule type: DNA
 A:Residues: 1-1298 <MCG>
 A:Cross-references: GB:X14112; GB:D00317; GB:D00374; GB:S40593; NID:g1944536; PIDN:CAA32
 C:Comment: This protein acts at the transcriptional regulatory level and is required for
 C:Genetics:
 A:Gene: IE3
 A:Map position: short repeat region (IR-s)
 C:Superfamily: herpesvirus immediate-early protein IE175
 C:Keywords: DNA binding; early protein; transcription regulation

Query Match 9.0%; Score 118.5; DB 1; Length 1298;
 Best Local Similarity 24.9%; Pred. No. 3;
 Matches 88; Conservative 24; Mismatches 121; Indels 121; Gaps 15;

QY 2 AGSP-LWGPAGVGLVLLLLGLFRPPPPALCARPVKEP----- 41
 Db 355 SGAPAAVAVPELGDAAQOYALITRLTPDAEAMGWLQNPVPGDVALDQACFRISGAA 414
 QY 42 -----GLSASPPLAETGAPRRFRSVPCEAGAVQ-----ELARAL 79
 Db 415 RNSSFTTGSVARAVPHLYAMAAGREGWGLAHAAAVAMSRDYDRAQKGLFLLTSRRAY 474
 QY 80 AHLEAER-----QERARAEAEEDQ-----ARVLAQLLRVWGAP 116
 Db 475 APLARENAAITGAGSFGAGADDEGVAVAAMAPGERAVPAGYGAAGILAAAGRLSAAP 534
 QY 117 RNSDPALGLDDDPAD-----AAQLAR-----ALLRARLDPAALAAQLVPAPV 159
 Db 535 --ASPAGG--DDPDAARHADADDAGRRAGQAGRVAVECLAACRCILEALAEFGDGLAAV 590
 QY 160 PA-AALRPPPVYDDGAPDAEAGDETPDVDPQLRLYLGR----- 202
 Db 591 PGLAGARPASPPRPGPAGP-----ASPPPHADAPRLRLRELFRVDAVLVLRGLD 646
 QY 203 -LAGSADSEGAAPRRLRRRAADHDVGSLEPPEGVLG-ALLRVKRLTETAPQVPA 254
 Db 647 RVAGGSEA-AVAARAVSLVA-----GALGPALPRDPRLPSSAAAAA 688

RESULT 13

A45344
 Immediate-early protein - suid herpesvirus 1 (strain Kaplan)
 C:Species: suid herpesvirus 1
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
 C:Accession: A45344
 R:Vicek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwzyer, M.
 Virology 179, 365-377, 1990
 A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented op

A:Reference number: A45344; MUID:91021039; PMID:2171211
 A:Accession: A45344
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1446 <VLC>
 A:Cross-references: GB:M34651; NID:g334070; PIDN:AAA47470.1; PID:g334071
 C:Superfamily: herpesvirus immediate-early protein IE175
 C:Keywords: DNA binding; early protein; transcription regulation

Query Match 8.9%; Score 118; DB 1; Length 1446;
 Best Local Similarity 26.6%; Pred. No. 3.6;
 Matches 83; Conservative 22; Mismatches 97; Indels 110; Gaps 15;

QY 16 GLLVLLLLGLFRPPPPALCARPVKEPGLSASAPPLAE-TGAPRRFRSVP-----RGEA 68
 Db 696 GVLERLL-----PCPLRLPAPARAPAAALGPAC--LEEVTAALLRALDAIPGAPERRQA 748
 QY 69 AGAVOELARALAHLEAERQERARAEAEQEDQOARVLQALLRVWVGAPRNSDPALGLDD 128
 Db 749 ADSVALVAVTAVPLV---RYSVDGARAREAAWTAAALF-----AFAN----- 788
 QY 129 PDAPAAQLARALLRARLDPAALAAQL-----VPAPVPAALRRPPPVYDDGPGAG 177
 Db 789 --VAAARLAE--AARCPAPAPCLPLWPEQPLVVPAPAPAAAGAP-SGLPCSGSPSS 843
 QY 178 PDAAEAGDET-----PDVDELLRY----- 197
 Db 844 PASTKSSSTKSSSTKSSGSGYASSPAAGPDPAPEPRKKRRAPGARPPGDEGE 903
 QY 198 -LLGRILAGSA---DSEGVAAAPRLRAADHDVGSLEPPEGVLGALL-----RV 242
 Db 904 GLSAAALRGDCHGRHDEEDRGPRKKRSLGLGAPADPAP-----ALLSSSSSSSEDDRL 958
 QY 243 KRLETPAPQVPA 254
 Db 959 RRLPGMPHEPA 970

RESULT 14

FOLJLK
 gag polyprotein - simian foamy virus (type 3, strain LK3)
 N:Alternate names: core polyprotein
 C:Species: simian foamy virus
 A:Note: host (African green monkey)
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 08-Apr-1994
 C:Accession: A40820
 R:Renne, R.; Friedl, E.; Schweizer, M.; Fleps, U.; Turek, R.; Neumann-Haefelin, D.
 Virology 186, 597-608, 1992
 A:Title: Genomic organization and expression of simian foamy virus type 3 (SFV-3).
 A:Reference number: A40820; MUID:92124734; PMID:1310187
 A:Accession: A40820
 A:Molecule type: DNA
 A:Residues: 1-643 <REN>
 A:Cross-references: GB:M74895
 C:Genetics:
 A:Gene: gag
 C:Superfamily: foamy virus gag polyprotein
 C:Keywords: core protein; polyprotein

Query Match 8.9%; Score 117; DB 1; Length 643;
 Best Local Similarity 22.2%; Pred. No. 1.8;
 Matches 66; Conservative 40; Mismatches 101; Indels 90; Gaps 14;

QY 8 WGPAGGVGLVLLGLFRPPPPALCARPVKEPGLSAAAP--PLAETGAPRRFRSVP 65
 Db 39 WGPCTR-----YLLVSIFLQDDSS--CQPLQOPRRWRPEGRPNVPLVINTI-----BAPW 84
 QY 66 GERAGAVOELARALAHLE-----LEAERQERARAEAEQEDQ 102
 Db 85 GELRQAFEDLDVABGTLRFGLANGWIPGDEYSMEFQPPLAQIAQMQORLELEIDIT 144
 QY 103 ARVLAQLL-----RWGAPRNSDDPALGLDD-----PDAPAAQLARALLRARL 145

[illegible]

Query Match	8.8%	Score	116.5	DB 2	Length	1677			
Best Local Similarity	25.9%	Pred. No.	5.2						
Matches	63	Conservative	26	Mismatches	113	Indels	41	Gaps	10
QY	28	PPFALCARPVKEPRLGISAAPP	LAETGAPRRFRSVP	RGAAAGVQE	-----	LARALAH	81		
Db	993	PPFSRAAGKVLPNGKVP	QAQPLQEARKKR	DI	SHAGQKGAAS	RGPEGTASPLGSGAPD	1052		
QY	82	LLBAERQER-ARAEAEAE	DOQARVLAQLLRV	GVGAPN	SDP-ALGDLDDDDP	AAQALARA	139		
Db	1053	LQESAMONLR	LATAEASLHQO	---VLSR----	HPOGSDP	VATSMFVQDVLQASTPATG	1103		
QY	140	LLRARLDPAALAAQLVP	PAPYAPPAALR	PPVPYDDG	PAGPDAEAGD	ETPDVDP	PELLRYLL	199	
Db	1104	VTQGSISAVAGSEAR	IPAVQKA	AVTDEP----	DHPTQGHQ	HEDSIQQAPEPLQPELLRIH	1160		
QY	200	GRILAGSADSEGA	VPARRLR	RAADHDV	GSPELVG	ALLVKVRL	ETPTAPQVPAR--RL	257	
Db	1161	NR-----	PSGOKTPE-----	GSETKPSKA	ESTML--PRKKPP	PPPPKPAHLSQI	1201		
QY	258	LPP	260						
Db	1202	HPP	1204						

Search completed: April 4, 2003, 14:18:38
Job time : 22 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:09:29 ; Search time 11 Seconds
(without alignments)
980.350 Million cell updates/sec

Title: US-09-803-589-6

Perfect score: 1319

Sequence: 1 MAGSPLLWCPAGGCVLLVL.....RVKRLTAPAPQVPARLLPP 260

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124.5	9.4	815	1 MK07_HUMAN	Q13164 homo sapien
2	123	9.3	1238	1 SBCC_RHOCA	O68032 rhodobacter
3	119.5	9.1	1461	1 IE18_PVIF	P11675 pseudobacter
4	118.5	9.0	1298	1 ICP4_HSV11	P08392 herpes simp
5	118	8.9	1446	1 IE18_PVKA	P33479 pseudorabie
6	117	8.9	643	1 GAG_SF3L	P27400 simian foam
7	116.5	8.8	616	1 VGF_HUMAN	O15240 homo sapien
8	116	8.8	548	1 ERF_HUMAN	P50548 homo sapien
9	116	8.8	551	1 ERF_MOUSE	P70459 mus musculus
10	113.5	8.6	806	1 MK07_MOUSE	Q9WV88 mus musculus
11	113.5	8.6	933	1 PRGR_HUMAN	P06401 homo sapien
12	113.5	8.6	3164	1 TEGU_HSV11	P10220 herpes simp
13	112	8.5	817	1 NEB2_RAT	O35274 rattus norv
14	111	8.4	2142	1 BAT2_HUMAN	P48634 homo sapien
15	110	8.3	1411	1 TCOF_HUMAN	Q13428 homo sapien
16	109.5	8.3	2564	1 SPCQ_HUMAN	Q9H254 homo sapien
17	108	8.2	319	1 FXE3_HUMAN	Q13461 homo sapien
18	107	8.1	676	1 ICP0_HSVBK	P29836 bovine herp
19	107	8.1	954	1 M3KA_HUMAN	Q02779 homo sapien
20	107	8.1	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r
21	106.5	8.1	395	1 TI22_HUMAN	Q9Y3q8 homo sapien
22	106	8.0	703	1 ML51_HUMAN	O15234 homo sapien
23	105	8.0	622	1 3BPI_HUMAN	Q9Y313 homo sapien
24	105	8.0	827	1 MOF_DROME	O02193 drosophila
25	105	8.0	1532	1 IGA_NEIGO	P09790 neisseria g
26	103.5	7.8	336	1 YB56_XENLA	P21574 xenopus lae
27	103.5	7.8	439	1 XP2_XENLA	P17437 xenopus lae
28	103.5	7.8	1184	1 PC12_HUMAN	Q9N994 homo sapien
29	103	7.8	603	1 NMBL_MOUSE	O08919 mus musculus
30	103	7.8	3301	1 CLR3_MOUSE	Q91210 mus musculus
31	102.5	7.8	553	1 ODO2_MYCTU	O10381 mycobacteri
32	102.5	7.8	628	1 V70K_TYWA	P20131 turnip yell
33	102.5	7.8	1402	1 IF4G_RABIT	P41110 oryctolagus

ALIGNMENTS

RESULT 1

ID	MK07_HUMAN	STANDARD;	PRT;	815 AA.
AC	Q13164; Q16634;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Mitogen-activated protein kinase 7 (EC 2.7.1.-) (Extracellular signal-regulated kinase 5) (ERK-5) (ERK4) (BMK1 Kinase).			
GN	MAPK7 OR PRK47 OR ERK5 OR ERK4.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=95279403; PubMed=7759517;			
RA	Zhou G., Bao Z.Q., Dixon J.E.;			
RT	"Components of a new human protein kinase signal transduction pathway.";			
RL	J. Biol. Chem. 270:12665-12669(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=95374539; PubMed=7646528;			
RA	Lee J.-D., Ulevitch R.J., Han J.;			
RT	"Primary structure of BMK1: a new mammalian map kinase.";			
RL	Biochem. Biophys. Res. Commun. 213:715-724(1995).			
CC	-1- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.			
CC	-1- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE PHOSPHORYLATION (BY SIMILARITY).			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE IN LIVER.			
CC	-1- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.			
CC	-1- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY ROLE, IS ABSENT.			
CC	-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. MAP KINASE SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U25278; AAA81381.1; -			
DR	EMBL; U25725; AAA82931.1; -			
DR	EMBL; U29726; AAA82932.1; -			
DR	EMBL; U29727; AAA82933.1; -			

Q9ukn7 homo sapien
P28284 herpes simp
Q9jlvl1 mus musculus
Q04637 homo sapien
P26651 homo sapien
P13627 paracoccus
P03972 bos taurus
Q96rd6 homo sapien
P29128 bovine herp
Q05152 oryctolagus
P74876 salmonella
Q9c0c2 homo sapien

OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31523;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89315207; PubMed=2546124;
RA Cheung A.K.;
RT "DNA nucleotide sequence analysis of the immediate-early gene of
RT pseudorabies virus.";
RL Nucleic Acids Res. 17:4637-4646(1989).
RN [2]
RN REVISIONS.
RN Cheung A.K.;
RA Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X15120; CAA33214.1;
CC PIR: S04713; EDBE7F.
CC InterPro: IPR005205; Herpes_ICP4_C.
CC InterPro: IPR005206; Herpes_ICP4_N.
CC Pfam: PF03584; Herpes_ICP4_N.
CC Pfam: PF03585; Herpes_ICP4_C.
CC Early protein: Transcription regulation; Trans-acting factor;
CC DNA-binding; Phosphorylation; Nuclear protein.
CC DOMAIN 390 405 POLY-SER.
CC DOMAIN 958 966 POLY-SER.
CC SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;
DR
DR EMBL: X15120; CAA33214.1;
DR PIR: S04713; EDBE7F.
DR InterPro: IPR005205; Herpes_ICP4_C.
DR InterPro: IPR005206; Herpes_ICP4_N.
DR Pfam: PF03584; Herpes_ICP4_N.
DR Pfam: PF03585; Herpes_ICP4_C.
KW Early protein: Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 390 405 POLY-SER.
FT DOMAIN 958 966 POLY-SER.
SQ SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 9.1%; Score 119.5; DB 1; Length 1461;
Best Local Similarity 25.1%; Pred. No. 1.9;
Matches 77; Conservative 27; Mismatches 105; Indels 98; Gaps 13;

QY 16 GLLVLLLLGLFRPPPPALCARPKVKEPRGLSASPPPLAE--TGAPRRFRRSVP-----RGEA 68
DB 707 GVLERLL-----PCPLRLPAPAPAPALGPAC--LEEVTAALLALRAIPGAGPAERQQA 759
QY 69 AGAVQELARALAHLEERQERARAEAOEADQOARVL-----AQLLRVWGAPRNSD 120
DB 760 ADSVALVARTVAPLV--RYSVDGARAREAAWTYAAALFAPANVAGARLAEEAARPGPAE 816
QY 121 PALGLDD-DPDAPAAQLARALLRARLDPALAAQLVAPVPAPAAALRPYPVDDPGAPD 179
DB 817 PAPGLPLMPQPGGL-----VVPAPAPAAAGAP-SGLPFGSGPSSPA 856
QY 180 AEEAGDET-----PDVDPPELLRY-----L 198
DB 857 STKSGSTKSSGKSGLSGSSGYASSPAAGPDPAPEPRKKRRAPGARPPGDGEDEGL 916
QY 199 LGRILASG-----DSEGVAPRRRLRAADHDVGSLELPEGVIGALL-----RVKRL 247
DB 917 SGSALRGDGHGRDDEEDRGPRKKRRSLGLGPADPDAPALVSSSSSSSSSSDRLRLRLP 976
QY 248 PAPQVPA 254
DB 977 PMPEHPA 983

RESULT 4
ICP4_HSV11

ICP4_HSV11 STANDARD; PRT; 1298 AA.
P08392;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Trans-acting transcriptional protein ICP4 (Transcriptional activator
DE IE175) (Alpha-4 protein).
GN ICP4 OR IE175 OR RSL.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1.";
RL J. Gen. Virol. 69:1531-1574(1988).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=86148504; PubMed=3005980;
RA McGeoch D.J., Dolan A., Donald S., Brauer D.H.K.;
RT "Complete DNA sequence of the short repeat region in the genome of
RT herpes simplex virus type 1.";
RL Nucleic Acids Res. 14:1727-1745(1986).
RN [3]
RN DNA-BINDING DOMAIN.
RX MEDLINE=90174974; PubMed=2155403;
RA Wu C.-L., Wilcox K.W.;
RT "Codons 262 to 490 from the herpes simplex virus ICP4 gene are
RT sufficient to encode a sequence-specific DNA binding protein.";
RL Nucleic Acids Res. 18:531-538(1990).
RN [4]
RN INFLUENCE OF PHOSPHORYLATION ON FUNCTION.
RX MEDLINE=9112047; PubMed=1846804;
RA Papavassiliou A.G., Wilcox K.W., Silverstein S.J.;
RT "The interaction of ICP4 with cell-infected-cell factors and its
RT state of phosphorylation modulate differential recognition of leader
RT sequences in herpes simplex virus DNA.";
RL EMBO J. 10:397-406(1991).
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS. BINDING
CC OF ICP4 TO ALPHA GENES CAUSES REPRESSION OF THEIR EXPRESSION.
CC ICP4 ALSO ACTS AS A TRANSCRIPTIONAL ACTIVATOR OF BETA AND GAMMA
CC GENES. ICP4 BINDS WITH HIGH AFFINITY TO THE SEQUENCE 5'-ATGCTC-3'.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -1- PTM: ICP4 IS ADP-RIBOSYLATED.
CC -1- PTM: THE LONG STRETCH OF SER IS A MAJOR SITE OF PHOSPHORYLATION.
CC ONLY THE PHOSPHORYLATED FORMS OF ICP4 ARE CAPABLE OF INTERACTING
CC WITH BETA OR GAMMA GENES.
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L00036; AAA96675.1;
CC EMBL: L00036; AAA96688.1;
CC EMBL: X14112; CAA32286.1;
CC EMBL: X14112; CAA32278.1;
CC EMBL: X06461; CAA29763.1;
CC PIR: A23510; EDBE75.
CC InterPro: IPR005205; Herpes_ICP4_C.
CC InterPro: IPR005206; Herpes_ICP4_N.
CC Pfam: PF03584; Herpes_ICP4_N;

```
DR Pfam: PF03585; Herpes_ICP4_C; 1.
KW Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein; ADP-ribosylation.
FT DOMAIN 176 199
FT DNA_BIND 262 490
FT SEQUENCE 1298 AA; 132843 MW; 4F32E04C95CA9344 CRC64;
SQ
Query Match 9.0%; Score 118.5; DB 1; Length 1298;
Best Local Similarity 24.9%; Pred. No. 1.9;
Matches 88; Conservative 24; Mismatches 121; Indels 121; Gaps 15;
QY 2 AGSP-LIAGPRAGGVLLVLLGLFRPPPPALCARPVKEPR----- 41
DB 355 SGAPAAVAPPELDAAQQVALITRLTYTPDAEMGWLQNPVRVPGVDALDQACFRISGAA 414
QY 42 -----GLSAASPPLAETGAPRRFRSVPGRGEAGAVQ-----ELARAL 79
DB 415 RNSSPFITGSVARVPHGLCYAAGFGWGLAHAAAVAMSRRYDRAQKGFLLTSLRAY 474
QY 80 AHLLAER-----QERARAEQAEEDQ-----ARVLAQLLRVWVGAP 116
DB 475 APLLARENAALTAAGSPGAGADDEGVAAVAAAAPGERAVPAGYGAAGTILAAALGRLSAAP 534
QY 117 RNSDPALGLDDPDAP-----AAQLAR-----ALLRLRDPALAAQLVPAPV 159
DB 535 --ASPAGG--DDPDAARHADADDAGRRRAQAGRVAVECUACRGILAEFGDGLAAV 590
QY 160 PA-AALRPFRPPVDDGPAGPADEAGDETPDVPDELLRYLLGRI 202
DB 591 PGLAGARPASPPRPEGAPG-----ASPPPHADAPRLRAWLRELFVRVDALVLMRLRGDL 646
QY 203 -LAGSADSEGVAAAPRLRRAADHDVGSSELPPEGVLG-ALLRVKRLTETAPQVPA 254
DB 647 RVAGGSEA-AVAARAVSLVA-----GALGPALPRDPRPLPSSAAAAA 688
RESULT 5
IE18_PVKKA
ID IE18_PVKKA STANDARD; PRT: 1446 AA.
AC P33479;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE Immediate-early protein IE180.
GN IE.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; PubMed=2171211;
RA Vilek C., Kozmik Z., Paces V., Schirm S., Schwytzer M.;
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT oriented open reading frame: characterization of their promoter and
RT enhancer regions.";
RL Virology 179:365-377(1990).
CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -!- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL: M34651; AAA47470.1; -.
DR PIR: A45344; A45344.
DR InterPro: IPR005205; Herpes_ICP4_C.
DR InterPro: IPR005206; Herpes_ICP4_N.
DR Pfam: PF03584; Herpes_ICP4_N; 1.
DR Pfam: PF03585; Herpes_ICP4_C; 1.
KW Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 347 354
FT DOMAIN 379 397
FT POLY-SER.
SQ SEQUENCE 1446 AA; 148640 MW; 81F43A3DE3DDA068 CRC64;
Query Match 8.9%; Score 118; DB 1; Length 1446;
Best Local Similarity 26.6%; Pred. No. 2.3;
Matches 83; Conservative 22; Mismatches 97; Indels 110; Gaps 15;
QY 16 GLIVLLLLGLFRPPPPALCARPVKEPRGLSASPPLAE-TGAPRRFRSVP-----RGEA 68
DB 696 GVLRL-----PCRLPAPAPAPAAALGPAC--LEEVTAALLALRDAIPGAGPAERROA 748
QY 69 AGAVOELARALAHLEAEQERARAEQAEQDQARVLAQLLRVWVGAPRNSDPALGLDDD 128
DB 749 ADSVALVARTVAPLV---RYSVDGARAREAAWTYAAALF-----APAN----- 788
QY 129 PDAPAAQLARALLRLRDLPAALAAQL-----VPAPVPAALRPPRPVYDDGPAG 177
DB 789 --VAARALAE--AARPGPAEPAGLPPLWPEQGLVVPAPAPAAAGAP-SGLPGSGPSS 843
QY 178 PDAAEAGDET-----POVDPELLRY----- 197
DB 844 PASTKSSSTKSSSTKSSGSSGYASSPAAGDPAPERRKKRRAPGARRPGDEE 903
QY 198 -LLGRILAGSA-----DSEGVAAPRLRRAADHDVGSSELPPEGVLGALL-----RV 242
DB 904 GLSGAALRGDGHGRDDEEDGRGRRKRSGLGLGAPDPAP-----ALLSSSSSSSEDDRL 958
QY 243 KRLETPAPQVPA 254
DB 959 RRLPGMPPEHPA 970
RESULT 6
GAG_SFV3L
ID GAG_SFV3L STANDARD; PRT: 643 AA.
AC P27400;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE GAG polyprotein (Core polyprotein).
GN GAG.
OS Simian foamy virus (type 3 / strain LK3) (SFV-3).
OC Viruses; Retroid viruses; Retroviridae; Spumavirus.
OX NCBI_TaxID=11644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9214734; PubMed=1310187;
RA Renne R., Friedl E., Schweizer M., Fleps U., Turek R.,
RA Neumann-Haefelin D.;
RT "Genomic organization and expression of simian foamy virus type 3
RT (SFV-3).";
RL Virology 186:597-608(1992).
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL: M74895; AAA47795.1; ALT_INIT.
DR PIR: A40820; FOLJLK.
```

DR InterPro: IPR004957; Spuma_gag.
DR Pfam: PF03276; Gag_spuma; 1.
KW Core protein; Polyprotein.

SQ SEQUENCE 643 AA; 69785 MW; C53A0575BA9B5949 CRC64;

Query Match 8.9%; Score 117; DB 1; Length 643;
Best Local Similarity 22.2%; Pred. No. 1.2; Indels 90; Gaps 14;
Matches 66; Conservative 40; Mismatches 101; Indels 90; Gaps 14;

QY 8 WGPAGVGLVLLGLLGRPPPPALCARPVKEPRGLSAA--PLAETGAPRRFRSVPR 65
DB 39 WGPCTR-----YLVSIPLQDSS--GQLQPPWRPEGRVPLVHTI-----EAPW 84
QY 66 GAGAGVQELARALAH-----LEAERQRAAEAEQEDQ 102
DB 85 GELQAFEDLDVAEGTLRFGLANGWIPGDEYSMEFQPPPLAQEIAQMORDEEILDI 144
QY 103 ARVLAQL-----RVMGAPRNSDPALGLDD-----PDAPAAQLARALLR 145
DB 145 GQICAQVIDLVMDQAIQIRLERIQRLGLRNLVAGIQAPSSPIGO----- 194
QY 146 DPAALAAQLVPAPVPAALPRPPVYDDGPGAPDAEAGDET--DVDPELLRYLLGLR 204
DB 195 -----PIASSLPPIP-----GSSSPADLDGIWTPRQIDPRLSRVAYNPF 237
QY 205 GSADSEGVAPRLRRRAADHDVSELP--PEGVLGALLR-VKRLTPAPO-VPARRL 257
DB 238 GSSDGGSGSIP--VQPSAPPVAVLSPLPSPVQSIQVVAQPPVPAQPIQI 292

RESULT 7
VGF_HUMAN

ID VGF_HUMAN STANDARD; PRT; 616 AA.
AC O15240;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurosecretory protein VGF precursor.
GN VGF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RX MEDLINE=98008940; PubMed=9344675;
RA Canu N., Possenti R., Ricco A.S., Rocchi M., Levi A.;
RT "Cloning, structural organization analysis and chromosomal assignment
of the human gene for neurosecretory protein VGF.";
RL Genomics 45:443-446(1997).
CC -!- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF CELL-CELL
INTERACTIONS OR IN SYNATOGNESIS DURING THE MATURATION OF THE
NERVOUS SYSTEM (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Stored in secretory vesicles and then
secreted (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in brain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; Y12661; CAAT3210.1; -;
CC Genew; HGNC:12684; VGF.
CC MIM; 602186; -;
KW Growth factor; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 616 NEUROSECRETORY PROTEIN VGF.
FT DOMAIN 353 447 ASP/GLU-RICH (ACIDIC).

SQ SEQUENCE 616 AA; 67286 MW; CD1920610201BBB9 CRC64;

Query Match 8.8%; Score 116.5; DB 1; Length 616;
Best Local Similarity 25.9%; Pred. No. 1.2;
Matches 69; Conservative 27; Mismatches 109; Indels 61; Gaps 13;

QY 11 RAGVGVLVLLGLLGLRPPPPALCARPVKEPRGLSAA--SPPLAETGAPRRFRSVPRGEA 68
DB 5 RLSASALFCILLGLGAAPP--GRPEAQPPPLSSEHKFVAGDAVP-----GPK 52
QY 69 AGAVQELARALAHLEAERQRAAEAEQEDQARVLAQ--LLRVWGAPRNSDPALGLDD 127
DB 53 DGSAPEVRGA-----RNSEFQDEGELEFGVDPRALAAVLQALDRPASPPAPSSQ 104
QY 128 DPAPAAQLARALLRRLDPAALAAQLVPAPVPAALPRPPVYDDGPGAPDAEAGDET 187
DB 105 GPEERAAE--ALLTETVRSQTHSLPAAGEPEAPAPRPQT--ENGP-----EASDPS 153
QY 188 PDVDPPELLRYLLGLRILAGSADSGVAPRLRRRAADHDVSELPPEGVLGALLRVKRL 247
DB 154 EELE-----ALASLLQELRDFSPSSAKRQOETA-----AETETRTTLTRV-NLES 199
QY 248 PAPO-----VPARRLLPP 260
DB 200 PGPVWRASWGEFQARVPERAPLPP 225

RESULT 8
ERF_HUMAN

ID ERF_HUMAN STANDARD; PRT; 548 AA.
AC P50548; Q9UP17;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ETS-domain transcription factor ERF (Ets2 repressor factor).
GN ERF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=96030784; PubMed=7598608;
RA Sgouras D.N., Athanasios M.A., Beal G.J. Jr., Fisher R.J., Blair D.G.,
RA Mavrothalassitis G.J.;
RT "ERF: an ETS domain protein with strong transcriptional repressor
activity, can suppress ets-associated tumorigenesis and is regulated
by phosphorylation during cell cycle and mitogenic stimulation.";
RL EMBO J. 14:4781-4793(1995).
RN [2]
SEQUENCE FROM N.A.
RP Lamerdin J.E., McGready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Dangnan L., Erier A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andreise T., Frankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 3.2 Mb region in 19q13.2 between CYP2F1 and
DL9S178.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POTENT TRANSCRIPTIONAL REPRESSOR THAT BINDS TO THE H1
ELEMENT OF THE ETS2 PROMOTER. MAY REGULATE OTHER GENES INVOLVED
IN CELLULAR PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: PHOSPHORYLATED BY MULTIPLE KINASES INCLUDING PROBABLY ERK2.
CC PHOSPHORYLATION REGULATES THE ACTIVITY OF ERF.
CC -!- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mitogen-activated protein kinase 7 (EC 2.7.1.-) (Extracellular signal-
 DE regulated kinase 5) (ERK-5) (BMK1 kinase).
 GN MAPK7 OR ERK5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kamakura S., Moriguchi T., Nishida E.;
 RT Activation of the protein kinase ERK5/BMK1 by receptor tyrosine
 RT kinases: identification and characterization of a signaling pathway to
 RT the nucleus.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND
 CC NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS (BY SIMILARITY).
 CC -1- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE
 CC PHOSPHORYLATION (BY SIMILARITY).
 CC -1- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN
 CC TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.
 CC -1- PTM: AUTOPHOSPHORYLATION ON THREONINE AND TYROSINE RESIDUES, WHEN
 CC THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY
 CC ROLE, IS ABSENT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MAP KINASE SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB019373; BAA82039.1; -
 DR HSSP; P24941; 1HCL.
 DR MGD; MGI:1346347; Mapk7.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR003527; MAP_kin.
 DR InterPro: IPR002230; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00220; S_TKC; 1.
 DR PROSITE; PS01351; MAPK; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
 DR Transferrase; Serine/threonine-protein kinase; ATP-binding; Cell cycle;
 KW Phosphorylation.
 KW Phosphorylation.
 FT DOMAIN 55 347 PROTEIN KINASE.
 FT DOMAIN 434 465 PRO-RICH 1.
 FT DOMAIN 521 524 POLY-ARG.
 FT DOMAIN 578 700 PRO-RICH 2.
 FT NP_BIND 61 69 ATP (BY SIMILARITY).
 FT BINDING 84 84 ATP (BY SIMILARITY).
 FT ACT_SITE 182 182 BY SIMILARITY.
 FT MOD_RES 219 219 PHOSPHORYLATION (ACTIVATES THE KINASE)
 FT (BY SIMILARITY).
 FT MOD_RES 221 221 PHOSPHORYLATION (ACTIVATES THE KINASE)
 FT (BY SIMILARITY).
 SQ SEQUENCE 806 AA; 87732 MW; E7CC41C4BDE0633 CRC64;
 Query Match 8.6%; Score 113.5; DB 1; Length 806;
 Best Local Similarity 22.8%; Pred. No. 2.5;
 Matches 70; Conservative 34; Mismatches 88; Indels 115; Gaps 16;
 QY 5 PLLWGRAGGVGLLVLLILGLFPPPAL--CARPVKEPGLS-AASPLAETGAPRRFR 61
 DB 420 PSFWASGD-----CAMSEFPFALPCSDPAPDVTDLTQPAPPAELAPPKR--- 467

QY 62 SVPRGEAAGVQELARAL--AHLEA-----ERQER 90
 DB 468 -----EGAISDNTKAALKSLRSLRDGSPASLEAPEPRKDVTAQERQREER 520
 QY 91 ARAEQAEDQOQARVLAQLLRVVCAPR----NSDPALGL---DDDPAPAAQLARALL-- 141
 DB 521 RRRQERAKERKRQERKRGAGTGLGGFSTDPLAGLVLSND-----RSLER 571
 QY 142 --RARLDPAALAAQLVPAPAAALRPVPVYDDGPAGDAEEAGDTPDVPDPELLRYLL 199
 DB 572 WTRWAREPPAP-----APAPAPAPAPSSAQPTSTPTGTPVSQSTGFLQP----- 615
 QY 200 GRILAGSAD---SEGVAAPRRRLRRADHDVGSSELPPEGVGLALLRVKRLTPAPQ----VP 253
 DB 616 ----AGSIPGASQPVCP-----PGVPQAPGPIAPLQT-----APSTSLLA 655
 QY 254 ARLLPP 260
 DB 656 QSLSVPP 662
 RESULT 11
 PRGR_HUMAN
 ID PCR_HUMAN STANDARD; PRT; 933 AA.
 AC P06401; Q9UPF7;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Progesterone receptor (PR).
 GN PR OR NR3C3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=90228361; PubMed=2328727;
 RA Kastner P., Krust A., Turcotte B., Stropp U., Tori L., Gronemeyer H.,
 RA Chambon P.;
 RT "Two distinct estrogen-regulated promoters generate transcripts
 RT encoding the two functionally different human progesterone receptor
 RT forms A and B.";
 RL EMBO J. 9:1603-1614(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87184565; PubMed=3551956;
 RA Misrahi M., Atger M., D'Auriol L., Loosfelt H., Meriel C.,
 RA Fridlansky F., Guiochon-Mantel A., Galibert F., Milgrom E.;
 RT "Complete amino acid sequence of the human progesterone receptor
 RT deduced from cloned cDNA.";
 RL Biochem. Biophys. Res. Commun. 143:740-748(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kieback D.G., Agoulrik I.U., Tong X.-W.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 682-933.
 RX MEDLINE=98282128; PubMed=9620806;
 RA Williams S.P., Sigler P.B.;
 RT "Atomic structure of progesterone complexed with its receptor.";
 RL Nature 393:392-396(1998).
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
 CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
 CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A AND B (SHOWN HERE); ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
 CC NR3 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

BAT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 RP TISSUE=Testis;
 RC MEDLINE=90192810; PubMed=2156268;
 RA Banerji J., Sands J., Strominger J.L., Spies T.;
 RT "A gene pair from the human major histocompatibility complex encodes
 RT large proline-rich proteins with multiple repeated motifs and a
 RT single ubiquitin-like domain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).
 RN [2]
 SEQUENCE OF 1-1860 FROM N.A.
 RP MEDLINE=93272029; PubMed=8499947;
 RA Iris F.J.M., Bouguieret L., Prieur S., Caterina D., Primas G.,
 RA Perrot V., Jurka J., Rodriguez-Tome P., Clavierie J.-M., Dausset J.,
 RA Cohen D.;
 RT "dense Alu clustering and a potential new member of the NF kappa B
 RT family within a 90 kilobase HLA class III segment.";
 RL Nat. Genet. 3:137-145(1993).
 CC -1- FUNCTION: UNKNOWN.
 CC -1- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR ENBL; M33509; AAA35585.1; -;
 DR ENBL; M33518; AAA35586.1; -;
 DR ENBL; M33512; AAA35586.1; JOINED.
 DR ENBL; Z15025; CAA78744.1; -;
 DR PIR; B35098; B35098.
 DR PIR; S36152; S36152.
 DR Genew; HGNC:13918; BAT2.
 DR MIM; 142580; -;
 KW Repeat.
 FT DOMAIN 519 524 POLY-PRO.
 FT DOMAIN 636 657 GLN-RICH.
 FT DOMAIN 684 688 POLY-PRO.
 FT DOMAIN 699 704 POLY-PRO.
 FT DOMAIN 814 821 POLY-PRO.
 FT DOMAIN 1340 1345 POLY-GLY.
 FT DOMAIN 1398 1403 POLY-GLY.
 FT DOMAIN 1436 1442 POLY-PRO.
 FT DOMAIN 1982 1991 POLY-PRO.
 FT DOMAIN 41 1795 4 X 57 AA TYPE A REPEATS.
 FT REPEAT 41 95 1-1.
 FT REPEAT 98 154 1-2.
 FT REPEAT 281 337 1-3.
 FT REPEAT 1740 1795 1-4.
 FT DOMAIN 337 549 2 X TYPE B REPEATS.
 FT REPEAT 337 418 2-1.
 FT REPEAT 476 549 2-2.
 FT DOMAIN 1899 2089 3 X 50 AA TYPE C REPEATS.
 FT REPEAT 1899 1948 3-1.
 FT REPEAT 1965 2014 3-2.
 FT REPEAT 2040 2089 3-3.
 FT CONFLICT 57 57 R -> A (IN REF. 2).
 FT CONFLICT 109 109 Q -> S (IN REF. 2).
 FT CONFLICT 414 414 P -> PPHRGPGAGNMGP (IN REF. 2).
 FT CONFLICT 532 532 T -> K (IN REF. 2).
 FT CONFLICT 682 682 Q -> K (IN REF. 2).
 FT CONFLICT 730 730 E -> D (IN REF. 2).
 FT CONFLICT 750 750 L -> R (IN REF. 2).
 FT CONFLICT 834 834 A -> T (IN REF. 2).
 FT CONFLICT 1035 1035 G -> A (IN REF. 2).
 FT

Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).

[4]

VARIANTS L-439; V-810; V-1313 AND G-1355, AND VARIANT TCS R-53.

MEDLINE-97195537; PubMed-9042910;

Edwards S.J., Gladwin A.J., Dixon M.J.;

"The mutational spectrum in Treacher Collins syndrome reveals a predominance of mutations that create a premature-termination codon.";

Am. J. Hum. Genet. 60:515-524(1997).

-1- DISEASE: DEFECTS IN TCOF1 ARE THE CAUSE OF TREACHER COLLINS SYNDROME (TCS). TCS IS A AUTOSOMAL DOMINANT DISORDER OF CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000 LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALLY SYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS, ARESIA OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS; (2) LATERAL DOWNWARD SLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND ZYGOMATIC COMPLEX; (4) CLEFT PALATE.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U40847; AAC50903.1; -

EMBL; U76366; AAC51181.1; -

EMBL; U84664; AAC51185.1; -

EMBL; U84640; AAC51185.1; JOINED.

EMBL; U84641; AAC51185.1; JOINED.

EMBL; U84642; AAC51185.1; JOINED.

EMBL; U84643; AAC51185.1; JOINED.

EMBL; U84644; AAC51185.1; JOINED.

EMBL; U84645; AAC51185.1; JOINED.

EMBL; U84646; AAC51185.1; JOINED.

EMBL; U84647; AAC51185.1; JOINED.

EMBL; U84648; AAC51185.1; JOINED.

EMBL; U84649; AAC51185.1; JOINED.

EMBL; U84650; AAC51185.1; JOINED.

EMBL; U84651; AAC51185.1; JOINED.

EMBL; U84652; AAC51185.1; JOINED.

EMBL; U84653; AAC51185.1; JOINED.

EMBL; U84654; AAC51185.1; JOINED.

EMBL; U84655; AAC51185.1; JOINED.

EMBL; U84656; AAC51185.1; JOINED.

EMBL; U84657; AAC51185.1; JOINED.

EMBL; U84658; AAC51185.1; JOINED.

EMBL; U84659; AAC51185.1; JOINED.

EMBL; U84660; AAC51185.1; JOINED.

EMBL; U84661; AAC51185.1; JOINED.

EMBL; U84662; AAC51185.1; JOINED.

EMBL; U84663; AAC51185.1; JOINED.

EMBL; U79659; AAB40722.1; -

EMBL; U79645; AAB40722.1; JOINED.

EMBL; U79646; AAB40722.1; JOINED.

EMBL; U79647; AAB40722.1; JOINED.

EMBL; U79648; AAB40722.1; JOINED.

EMBL; U79649; AAB40722.1; JOINED.

EMBL; U79650; AAB40722.1; JOINED.

EMBL; U79651; AAB40722.1; JOINED.

EMBL; U79652; AAB40722.1; JOINED.

EMBL; U79653; AAB40722.1; JOINED.

EMBL; U79654; AAB40722.1; JOINED.

EMBL; U79655; AAB40722.1; JOINED.

EMBL; U79656; AAB40722.1; JOINED.

EMBL; U79657; AAB40722.1; JOINED.

EMBL; U79658; AAB40722.1; JOINED.

Genew; HGNC:11654; TCOF1.

MIN; 606847; -

MIN; 154500; -

DR InterPro: IPR003993; treacle.

DR Pfam: PF03546; treacle; 3.

DR PRINTS; PRO1503; TREACLE.

KW Disease mutation; Polymorphism.

FT DOMAIN 89 97 POLY-GLU.

FT DOMAIN 204 207 POLY-SER.

FT DOMAIN 516 519 POLY-SER.

FT DOMAIN 919 924 POLY-SER.

FT DOMAIN 1285 1289 POLY-LYS.

FT DOMAIN 1375 1386 POLY-LYS.

FT DOMAIN 1398 1405 POLY-LYS.

FT VARIANT 53 53 W -> R (IN TCS).

FT /FTid=VAR_005630.

FT P -> L.

FT VARIANT 439 439 /FTid=VAR_005631.

FT A -> V.

FT VARIANT 810 810 /FTid=VAR_005632.

FT A -> V.

FT VARIANT 1313 1313 /FTid=VAR_005633.

FT D -> G.

FT /FTid=VAR_005634.

FT K -> Q (IN REF. 2).

FT CONFLICT 1312 1312

SQ SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;

Query Match 8.38; Score 110; DB 1; Length 1411;

Best Local Similarity 21.5%; Pred. No. 7.1;

Matches 56; Conservative 37; Mismatches 80; Indels 88; Gaps 10;

QY 43 LSAASPLAETGAPRRFRSVPGEAGAGVQELARALAHLLLEAEERARAEQAEQDQ- 101

DB 294 VQAASAPAKE--SPKGAAPAPPGKTGPV---AKAQA---GKREDSQSSEESDSEE 344

QY 102 ----QARV--LAQLRVWGAPRNSDPALGLDDDP---DAPAA----- 134

DB 345 EPAQAKSPGKAPQVRAASAPAKESPRKGAAPAPPRKTGPAAQVQVKGQEDSRSSSEE 404

QY 135 -----QLARALLRLDPAALAAQLVPA-----PVPAAALRPPTVDDGPA 176

DB 405 SDSREALANMAAQVKPLGKSPQVKPASTMGMPGLKGAGFVPPGKYGPATPSAQVGKW 464

QY 177 GPDAEEAGDETPDVPDELLRYLLGRILAGSADSEGAAPRRLRRAADHDVGSELPP--EG 234

DB 465 EEDSESSSEESSD-----SSDGEVPTAVAPAEK 493

QY 235 VLGALLRVKRLTETAPQVPPAR 255

DB 494 SLGNILQAKPTSSPAKGPQK 514

Search completed: April 4, 2003, 14:16:41

Job time : 17 secs

Handwritten scribble or mark.

Vertical handwritten marks or characters.

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:15:40 ; Search time 15 Seconds
(without alignments)
509.997 Million cell updates/sec

Title: US-09-803-589-6

Perfect score: 1319

Sequence: 1 MAGSLLNGPRAGGVGLLV.....RVKRLTEPAPQVPRRLPPP 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.AA.*
1: /cgn2.6/prodata/1/iaa/5A_COMB.pep.*
2: /cgn2.6/prodata/1/iaa/5B_COMB.pep.*
3: /cgn2.6/prodata/1/iaa/6A_COMB.pep.*
4: /cgn2.6/prodata/1/iaa/6B_COMB.pep.*
5: /cgn2.6/prodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2.6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137.5	10.4	882	4	US-09-413-814-78
2	119.5	9.1	248	3	US-08-341-018-52
3	119.5	9.1	248	4	US-08-470-335-210
4	119.5	9.1	248	4	US-08-470-339-210
5	119.5	9.1	248	4	US-08-467-602-207
6	119.5	9.1	248	4	US-08-467-602-404
7	119.5	9.1	349	4	US-08-470-335-188
8	119.5	9.1	382	4	US-08-467-602-382
9	119.5	9.1	405	4	US-08-467-602-384
10	119.5	9.1	411	4	US-08-470-339-189
11	119.5	9.1	414	4	US-08-470-339-188
12	119.5	9.1	422	1	US-08-036-555B-170
13	119.5	9.1	422	1	US-08-469-569-170
14	119.5	9.1	422	1	US-08-428-926-3
15	119.5	9.1	422	1	US-08-249-322A-170
16	119.5	9.1	422	1	US-08-428-927-3
17	119.5	9.1	422	1	US-08-428-298-3
18	119.5	9.1	422	1	US-08-339-517-3
19	119.5	9.1	422	1	US-08-469-526A-170
20	119.5	9.1	422	2	US-08-734-591A-170
21	119.5	9.1	422	2	US-08-469-660-170
22	119.5	9.1	422	3	US-08-341-018-72
23	119.5	9.1	422	4	US-08-470-335-170
24	119.5	9.1	422	4	US-08-735-021-170
25	119.5	9.1	422	4	US-08-734-664A-170
26	119.5	9.1	422	4	US-08-470-339-170
27	119.5	9.1	422	4	US-08-467-602-170

28	119.5	9.1	422	4	US-08-467-602-324	Sequence 324, App
29	119.5	9.1	422	5	PCT-US94-05083C-166	Sequence 166, App
30	119.5	9.1	422	5	PCT-US94-05083C-185	Sequence 185, App
31	119.5	9.1	422	5	PCT-US95-06846A-170	Sequence 170, App
32	119.5	9.1	425	4	US-08-470-335-226	Sequence 226, App
33	119.5	9.1	425	4	US-08-467-602-320	Sequence 320, App
34	119.5	9.1	445	4	US-08-467-602-328	Sequence 328, App
35	119.5	9.1	456	4	US-08-470-335-246	Sequence 246, App
36	119.5	9.1	456	4	US-08-467-602-303	Sequence 303, App
37	119.5	9.1	456	4	US-08-467-602-366	Sequence 366, App
38	119.5	9.1	459	4	US-08-470-335-239	Sequence 239, App
39	119.5	9.1	459	4	US-08-467-602-299	Sequence 299, App
40	119.5	9.1	459	4	US-08-467-602-362	Sequence 362, App
41	119.5	9.1	479	4	US-08-467-602-307	Sequence 307, App
42	119.5	9.1	479	4	US-08-467-602-370	Sequence 370, App
43	119.5	9.1	490	4	US-08-467-602-345	Sequence 345, App
44	119.5	9.1	493	4	US-08-467-602-341	Sequence 341, App
45	119.5	9.1	513	4	US-08-467-602-349	Sequence 349, App

ALIGNMENTS

RESULT 1
US-09-413-814-78
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-78

Query Match 10.4%; Score 137.5; DB 4; Length 882;
Best Local Similarity 27.2%; Pred. No. 0.00031;
Matches 71; Conservative 13; Mismatches 72; Indels 105; Gaps 12;
QY 9 GRAGGVGLLVLLGLFRPPPCALCARPVKEPRGL-----SAASPPLAETGAPRRFRSVP 64
Db 559 GPRGGG-----APRGEPRVPRPEPPRAGHPPEAPRRHAR 601
QY 65 RGEAGAVQELARALAHLLAEERQARABAEQOQARVLAQLLRVWGAPRNSDPALG 124
Db 602 RA-----PRVRRLVGRRLRRAR-----BALRRL-----RAGPAF- 632
QY 125 LDDDDPAPAAQARALLARLDPAALAAQLVPAPVPAALRPPPVYDD---GPAGPDAE 181
Db 633 -----PAAGCAPCAVRRLRRSPAGVAVRRGP-----GRAPVLDDALGRAAGPR 677
QY 182 EAGDETPDVDPPELLRYLLGRILAGSAD--SEGVAAPRRLRRRAADHD-----VGSLEPPEG 234
Db 678 -----AGRPRAGGAKDVARGAAEPRRRGRRAHPPDQGVLRAGGDLPLHG 720

QY 235 VLGALLRVKRLTETAPQVPAR 255
Db 721 -----PARRVOAR 728

RESULT 2

US-08-341-018-52
; Sequence 52, Application US/08341018A
; Patent No. 6087323
; GENERAL INFORMATION:
; APPLICANT: Gwynne, David I.
; APPLICANT: Mahanthappa, Nagesh K.
; APPLICANT: Marchionni, Mark A.
; APPLICANT: Bermingham-McDonogh, Olivia
; APPLICANT: Goldin, Stanley M.
; APPLICANT: McBurney, Robert N.
; TITLE OF INVENTION: USE OF NEUREGULINS AS MODULATORS OF
; TITLE OF INVENTION: CELLULAR COMMUNICATION
; FILE REFERENCE: 04585/041001
; CURRENT APPLICATION NUMBER: US/08/341.018A
; CURRENT FILING DATE: 1994-11-17
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-341-018-52

Query Match 9.1%; Score 119.5; DB 3; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;
QY 2 AGSPLLWGPAGGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRQRPASAARSSPPLPLLLLLGLTAAALPAGAAAGNEAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAGAVQELARALAHLEAEQERARAEQAEEDQQARVLAQ 108
Db 67 P-----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAG 106
QY 109 LLRVGAPRNSDPALGLDDDPAPAAQLARALLRRLDPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGAPGDAEEAGDETP 188
Db 141 PSWPTAPV-PSAGEGGEAP 159

RESULT 3

US-08-470-335-210
; Sequence 210, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROOBANT, PAUL
; APPLICANT: MINGHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470.335F
; CURRENT FILING DATE: 1993-06-06
; EARLIER APPLICATION NUMBER: 08/036,555
; EARLIER FILING DATE: 1993-03-24
; NUMBER OF SEQ ID NOS: 252
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210

; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-470-335-210

Query Match 9.1%; Score 119.5; DB 4; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;
QY 2 AGSPLLWGPAGGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRQRPASAARSSPPLPLLLLLGLTAAALPAGAAAGNEAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAGAVQELARALAHLEAEQERARAEQAEEDQQARVLAQ 108
Db 67 P-----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAG 106
QY 109 LLRVGAPRNSDPALGLDDDPAPAAQLARALLRRLDPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGAPGDAEEAGDETP 188
Db 141 PSWPTAPV-PSAGEGGEAP 159

RESULT 4

US-08-470-339-210
; Sequence 210, Application US/08470339C
; Patent No. 6232286
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROOBANT, PAUL
; APPLICANT: MINGHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470.339C
; CURRENT FILING DATE: 1993-06-06
; EARLIER APPLICATION NUMBER: 08/036,555
; EARLIER FILING DATE: 1993-03-24
; EARLIER APPLICATION NUMBER: 07/940,389
; EARLIER FILING DATE: 1992-09-03
; EARLIER APPLICATION NUMBER: 07/907,138
; EARLIER FILING DATE: 1992-06-30
; EARLIER APPLICATION NUMBER: 07/863,703
; EARLIER FILING DATE: 1992-04-03
; EARLIER APPLICATION NUMBER: 91 07566.3 GB
; EARLIER FILING DATE: 1999-04-10
; NUMBER OF SEQ ID NOS: 226
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-470-339-210

Query Match 9.1%; Score 119.5; DB 4; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;
QY 2 AGSPLLWGPAGGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRQRPASAARSSPPLPLLLLLGLTAAALPAGAAAGNEAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAGAVQELARALAHLEAEQERARAEQAEEDQQARVLAQ 108
Db 67 P-----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAG 106

```
QY 109 LLRWGAPRNSDPALGLDDPDPAQAQALRALRLRDLPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159
```

RESULT 5

```
US-08-467-602-207
; Sequence 207, Application US/08467602C
; Patent No. 6444642
; GENERAL INFORMATION:
; APPLICANT: Sklar, Robert
; APPLICANT: Marchionni, Mark
; APPLICANT: Gwynne, David I.
; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 04585/028003
; CURRENT APPLICATION NUMBER: US/08/467,602C
; CURRENT FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/209,204
; EARLIER FILING DATE: 1994-03-08
; EARLIER APPLICATION NUMBER: 08/059,022
; EARLIER FILING DATE: 1993-05-06
; NUMBER OF SEQ ID NOS: 420
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 207
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-467-602-207
```

```
Query Match 9.1%; Score 119.5; DB 4; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;
```

```
QY 2 AGSPLLWGPRAGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRAPRGSAAARSPPLPLPLLLLLGTAALAPGAAAGNEAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAAGAVQELARALAHLEAERERARAEAEQEDQOARVLAQ 108
Db 67 P-----SVGSVQELAQRAAANVIEGKVHPQR--QQGALDRKAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDPAQAQALRALRLRDLPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159
```

RESULT 6

```
US-08-467-602-404
; Sequence 404, Application US/08467602C
; Patent No. 6444642
; GENERAL INFORMATION:
; APPLICANT: Sklar, Robert
; APPLICANT: Marchionni, Mark
; APPLICANT: Gwynne, David I.
; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 04585/028003
; CURRENT APPLICATION NUMBER: US/08/467,602C
; CURRENT FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/209,204
; EARLIER FILING DATE: 1994-03-08
; EARLIER APPLICATION NUMBER: 08/059,022
; EARLIER FILING DATE: 1993-05-06
; NUMBER OF SEQ ID NOS: 420
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 404
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-467-602-404
```

```
Query Match 9.1%; Score 119.5; DB 4; Length 248;
Best Local Similarity 28.0%; Pred. No. 0.0026;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;
```

```
QY 2 AGSPLLWGPRAGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRAPRGSAAARSPPLPLPLLLLLGTAALAPGAAAGNEAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAAGAVQELARALAHLEAERERARAEAEQEDQOARVLAQ 108
Db 67 P-----SVGSVQELAQRAAANVIEGKVHPQR--QQGALDRKAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDPAQAQALRALRLRDLPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159
```

RESULT 7

```
US-08-470-335-188
; Sequence 188, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROEBANT, PAUL
; APPLICANT: MINCHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470,335F
; CURRENT FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/036,555
; EARLIER FILING DATE: 1993-03-24
; NUMBER OF SEQ ID NOS: 252
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-470-335-188
```

```
Query Match 9.1%; Score 119.5; DB 4; Length 349;
Best Local Similarity 28.0%; Pred. No. 0.004;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;
```

```
QY 2 AGSPLLWGPRAGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRAPRGSAAARSPPLPLPLLLLLGTAALAPGAAAGNEAPAGASVCYSSP 66
QY 49 PLAETGAPRRFRSVPGEAAGAVQELARALAHLEAERERARAEAEQEDQOARVLAQ 108
Db 67 P-----SVGSVQELAQRAAANVIEGKVHPQR--QQGALDRKAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDPAQAQALRALRLRDLPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
```


QY 49 PLAETGAPRRFRSVPGEAGAVOELARALAHLLLEAERQERARAEAEQEDQQARVLAQ 108
Db 67 P-----SVGSVOELAQRAAVVIEGKVHPQRR--QQGALDRKAAAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 11

US-08-470-339-188

; Sequence 188, Application US/08470339C

; Patent No. 6232286

; GENERAL INFORMATION:

; APPLICANT: GOODEARL, ANDREW

; APPLICANT: STROOBANT, PAUL

; APPLICANT: MINGHETTI, LUISA

; APPLICANT: WATERFIELD, MICHAEL

; APPLICANT: MARCHIONNI, MARK

; APPLICANT: CHEN, MARIO S.

; APPLICANT: HILES, IAN

; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR

; FILE REFERENCE: 04585/002008

; CURRENT APPLICATION NUMBER: US/08/470,339C

; EARLIER FILING DATE: 1995-06-06

; EARLIER FILING DATE: 1993-03-24

; EARLIER FILING DATE: 1992-09-03

; EARLIER FILING DATE: 1992-06-30

; EARLIER FILING DATE: 1992-04-03

; EARLIER FILING DATE: 1999-04-10

; NUMBER OF SEQ ID NOS: 226

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 188

; LENGTH: 414

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-470-339-188

Query Match 9.1%; Score 119.5; DB 4; Length 414;

Best Local Similarity 28.0%; Pred. No. 0.0049;

Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;

QY 2 AGSPLLWGPRAAGVG-----LLVLILLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRAPRGSARSPPPLPLPLLLLTALAPGAAAGNAAAGASVCYSPP 66
QY 49 PLAETGAPRRFRSVPGEAGAVOELARALAHLLLEAERQERARAEAEQEDQQARVLAQ 108
Db 67 P-----SVGSVOELAQRAAVVIEGKVHPQRR--QQGALDRKAAAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 12

US-08-036-555B-170

; Sequence 170, Application US/08036555B

; Patent No. 5530109

; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchionni, Mark;
; APPLICANT: Chen, Mario Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/036,555B

FILING DATE: 24-MAR-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/965,173

FILING DATE: 23-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/940,389

FILING DATE: 03-SEP-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/907,138

FILING DATE: 30-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/863,703

FILING DATE: 03-APRIL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: U.K. 91 07566.3

FILING DATE: 10-APRIL-1991

ATTORNEY/AGENT INFORMATION:

NAME: Tsai, Christine H.

REGISTRATION NUMBER: 34,266

REFERENCE/DOCKET NUMBER: LUD 5250.4

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 170:

SEQUENCE CHARACTERISTICS:

LENGTH: 422

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-08-036-555B-170

Query Match 9.1%; Score 119.5; DB 1; Length 422;

Best Local Similarity 28.0%; Pred. No. 0.0051;

Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps 9;

QY 2 AGSPLLWGPRAAGVG-----LLVLILLGLFRPPPPALCARPVKEPRGLSA--ASP 48
Db 10 SGRP---GPRAPRGSARSPPPLPLPLLLLTALAPGAAAGNAAAGASVCYSPP 66
QY 49 PLAETGAPRRFRSVPGEAGAVOELARALAHLLLEAERQERARAEAEQEDQQARVLAQ 108
Db 67 P-----SVGSVOELAQRAAVVIEGKVHPQRR--QQGALDRKAAAAAG 106
QY 109 LLRWGAPRNSDPALGLDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPRP 168
Db 107 EAGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPGAPDAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159

```

Db 107 EGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPAGPDAAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 14
US-08-428-926-3
; Sequence 3, Application US/08428926
; Patent No. 5667780
; GENERAL INFORMATION:
; APPLICANT: Ho, Wei-Hsien
; APPLICANT: Osheroff, Phyllis L.
; TITLE OF INVENTION: SENSORY AND MOTOR NEURON DERIVED FACTOR (SMDF)
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,926
; FILING DATE: 25-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/339517
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 853D4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-428-926-3
; Query Match 9.1%; Score 119.5; DB 1; Length 422;
; Best Local Similarity 28.0%; Pred. No. 0.0051;
; Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps

QY 2 AGSPLWGPAGVG-----LLVLLLLGLFRPPALCARPVPKEPGLSA--ASP 48
Db 10 SCRP---GPRAPRGSAARSSPPLPLLLGLTAAALPGAAAGNEAAPAGASVCVYSSP 66
QY 49 PLAETCAPRRFRKRVPRGEAGAVCELARALAHILLEAESRQERARAEQAEADQCARVLAQ 108
Db 67 P-----SVGSVELAQRAAVVIEGKVHQR--QQGALDRKAAAAG 106
QY 109 LRVWGAPRNSDPALGLDDDPDAPAAQLARALLRARLDPAALAAQLVPAPVPAALRPRP 168
Db 107 EGAWGG-----DREPPAA-----GPRALGP---PAEPLLAANGTV 140
QY 169 PVYDDGPAGPDAAEAGDETP 188
Db 141 PSWPTAPV-PSAGEPGEAP 159

RESULT 15

```

```

RESULT 13
US-08-469-569-170
; Sequence 170, Application US/08469569
; Patent No. 5606032
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchionni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,569
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 5250.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 170:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-469-569-170
Query Match 9.1%; Score 119.5; DB 1; Length 422;
Best Local Similarity 28.0%; Pred. No. 0.0051;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps
QY 2 AGSPLLWGPRAGVG-----LLVLLLLGLFRPPPPALCARPVKEPRGLSA--ASP 48
:| | | | | |
Ddb 10 SGRP--GPRAPQRPGSAARSSPPLPLLLILLGLTAAALPGAAAGNPAAPAGASVCYSSP 66
:| | | | | |
QY 49 PLAETGAPRRFRSVPGRGAAGVQELARALAHILLEAFERQERARAEAEQDOQARVLAQ 108
:| | | | | |
Ddb 67 P-----SVGSVELAQRAAVTEGVKHPQRR--QOGLDRKKAANAAG 106
:| | | | | |
QY 109 LLRWVGAPRNSDPALGLDDDDPAPAAQARALLRARLDPPAALAAQLVPAPVPAALPRP 168
:| | | | | |

```

```

US-08-249-322A-170
; Sequence 170, Application US/08249322A
; Patent No. 5716930
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Mao Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; NUMBER OF INVENTIONS: Preparation and Use
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/249,322A
; FILING DATE: 26-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 250.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 170:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-249-322A-170

Query Match          9.1%; Score 119.5; DB 1; Length 422;
Best Local similarity 28.0%; Pred. No. 0.0051;
Matches 56; Conservative 15; Mismatches 66; Indels 63; Gaps

QY      2   ACSPLLWGRAGVG-----LLVLLLLGLFRPPPCALCARPKVEPRGLSA--ASP 48
       :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      10  SGRP----GPAQRPGSAARSSPPLPLPLLTLTGTAALPGAAGNEAAPAGASYCYS 66
       :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY     49  PLAETGAPRRFRSVPRGEAGAVQLARALAHLEAEQERARAEAQEAEDQQARVLQA 108
       :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      67  P-----SVGSVQLAGRVAVVIECKVHPQRR--QQGALLDKKAAAAG 106
       ||
QY     109 LLRWGAPRNSDPALGLDDPDPAAPAQLARLRLARLDPAALAAQIVLPVPVPAALRP 168
       ||

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:18:20 ; Search time 36 Seconds
(without alignments)
441.538 Million cell updates/sec

Title: US-09-803-589-6

Perfect score: 1319

Sequence: 1 MAGSPLWGPAGGVGLLV.....RVKRLTPAPQVPRLLPP 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1319	100.0	260	10	US-09-803-589-6
2	627	47.5	175	10	US-09-803-589-14
3	308	23.4	149	9	US-09-984-271-150
4	308	23.4	166	9	US-09-984-271-238
5	124.5	9.4	418	9	US-09-946-807-3
6	124.5	9.4	418	10	US-09-795-668-3
7	124.5	9.4	418	10	US-09-795-686-3
8	118.5	9.0	550	9	US-09-976-740-47
9	118.5	9.0	550	12	US-10-023-529-47
10	118.5	9.0	550	12	US-10-023-523-47
11	118.5	9.0	1298	9	US-09-825-288A-2
12	114.5	8.7	492	10	US-09-764-898-206
13	112	8.5	2630	9	US-10-077-130-2
14	112	8.5	7968	9	US-10-077-130-5
15	110.5	8.4	538	9	US-09-976-740-43
16	110.5	8.4	538	12	US-10-023-529-43
17	110.5	8.4	538	12	US-10-023-523-43
18	109	8.3	681	10	US-09-815-242-11930
19	109	8.3	1665	10	US-09-858-664A-2

Sequence 3, Appli
Sequence 216, App
Sequence 216, App
Sequence 5, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 1037, Ap
Sequence 283, App
Sequence 100, App
Sequence 333, App
Sequence 331, App
Sequence 3, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 6, Appli
Sequence 8, Appli
Sequence 425, App
Sequence 425, App
Sequence 425, App
Sequence 425, App
Sequence 394, App
Sequence 394, App
Sequence 394, App
Sequence 394, App
Sequence 394, App
Sequence 394, App

20 108 8.2 846 10 US-09-858-664A-3
21 107.5 8.2 527 9 US-09-854-133-216
22 107.5 8.2 527 10 US-09-738-973-216
23 106 8.0 7257 9 US-10-014-717-5
24 105.5 8.0 1213 10 US-09-969-515-8
25 105.5 8.0 1224 10 US-09-969-515-2
26 105 8.0 404 10 US-09-764-864-1037
27 105 8.0 439 9 US-10-097-065-283
28 105 8.0 578 10 US-09-888-615-100
29 104 7.9 649 9 US-09-759-1308-333
30 104 7.9 671 9 US-09-759-1308-331
31 104 7.9 827 9 US-10-171-384-3
32 103.5 7.8 811 10 US-09-768-436-2
33 103.5 7.8 965 10 US-09-737-149-4
34 103.5 7.8 971 10 US-09-737-149-6
35 103.5 7.8 1043 10 US-09-737-149-8
36 103.5 7.8 1184 9 US-09-978-295A-425
37 103.5 7.8 1184 9 US-09-978-697-425
38 103.5 7.8 1184 9 US-09-978-192A-425
39 103.5 7.8 1184 9 US-09-999-832A-425
40 103.5 7.8 1184 9 US-09-978-189-425
41 103.5 7.8 1184 9 US-10-028-072-394
42 103.5 7.8 1184 9 US-10-121-049-394
43 103.5 7.8 1184 9 US-10-123-904-394
44 103.5 7.8 1184 9 US-10-140-470-394
45 103.5 7.8 1184 9 US-10-175-746-394

ALIGNMENTS

RESULT 1
US-09-803-589-6
; Sequence 6, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gooden, Andrew D.J.

; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: US6
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-803-589-6

Query Match 100.0%; Score 1319; DB 10; Length 260;
Best Local Similarity 100.0%; Pred. No. 9.9e-87;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLWGPAGGVGLLVLLLLLGLFRPPPPALCARPVKEPGLSAASPPLAETGAPRRFR 60
|||||

```
Db 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60
QY 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEQARVLAQLLRVWGAPRNSD 120
Db 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEQARVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPAPAAQARALLRRALDPAALAAQLVPAAPVPAALRRPPVPYDDGAPGPD 180
Db 121 PALGLDDDDPAPAAQARALLRRALDPAALAAQLVPAAPVPAALRRPPVPYDDGAPGPD 180
QY 181 EAGDFTPDVDPBELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPPBEGVLGALL 240
Db 181 EAGDFTPDVDPBELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPPBEGVLGALL 240
QY 241 RVKRLTAPQVPARRLLPP 260
Db 241 RVKRLTAPQVPARRLLPP 260

RESULT 2
US-09-803-589-14
; Sequence 14, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-803-589-14

Query Match 47.5%; Score 627; DB 10; Length 175;
Best Local Similarity 78.0%; Pred. No. 8.6e-38;
Matches 131; Conservative 6; Mismatches 27; Indels 4; Gaps 1;

QY 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60
Db 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASAPLVETSTPLRLR 60
QY 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEQARVLAQLLRVWGAPRNSD 120
Db 61 RAVPRGEAAGAVQELARALAHLEAEQERARAEAEQEQARVLAQLLRWAGSPRASD 120
QY 121 PALGLDDDDPAPAAQARALLRRALDPAALAAQLVPAAPVPAALRRPPVPYDDGAPGPD 164
Db 121 PPLAPDDDDPAPAAQARALLRRALDPAALAAQLVPAAPVPAALRRPPVPYDDGAPGPD 168

RESULT 3
```

```
US-09-984-271-150
; Sequence 150, Application US/09984271
; Publication No. US20030040088A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/984,271
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/482,273
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US99/15849
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: 60/092,921
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,922
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,956
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 150
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: xaa equals stop translation
US-09-984-271-150

Query Match 23.4%; Score 308; DB 9; Length 149;
Best Local Similarity 50.3%; Pred. No. 3.2e-15;
Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;

QY 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG 54
Db 1 MAGSPLWGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG 60
QY 55 --APRRFRSVPR-GEAAGAVQELARALAHLEAEQERARAEAEQEQARVLAQLLR 111
Db 61 GQCPVRRRGRCRPGAGAGA-----SAGAERQERARAEQOR-----LR 98
QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRARLDPAALAAQLVPA 157
Db 99 ISRRASW-----RSCCAGS-----APPATLIRLWANTTTTTLRQSSALCSAPA 143

RESULT 4
US-09-984-271-238
; Sequence 238, Application US/09984271
; Publication No. US20030040088A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/984,271
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/482,273
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US99/15849
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: 60/092,921
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,922
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,956
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 238
; LENGTH: 166
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-09-984-271-238

Query Match      23.4%; Score 308; DB 9; Length 166;
Best Local Similarity 50.3%; Pred. No. 3.6e-15;
Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;

QY 1 MAGSPILWGPAGVGGLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPLAETG-----54
DQ 19 MAGSPILWGPAGVGGLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPLARLALLAASG 78
QY 55 --APRRFRFRSVPR-GEAAGAVOELARALAHLLAEQERARAEAEQDQARVLAQLLR 111
DQ 79 GQCEVRRRCRGCGAGAGA-----SAGAEQERARAEQAR-----LR 116
QY 112 V-----WGAPRNSDPALGLDDDDPDAPAAQLAR-----ALLRARLDPAALAAQLVPA 157
DQ 117 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTPTLRQSSSLALCSAPA 161

RESULT 5
US-09-946-807-3
; Sequence 3, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-946-807-3

Query Match      9.4%; Score 124.5; DB 9; Length 418;
Best Local Similarity 28.8%; Pred. No. 0.11;
Matches 57; Conservative 14; Mismatches 64; Indels 63; Gaps 9;

QY 2 AGSPILWGPAGVG-----LLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPL 50
DQ 10 SGRP-----GPRARPGSAARSPPPLPLLLLLGLTAAAPGAA--AGNEAAPAGASVCSP- 63
QY 51 AETGAPRRFRSVPRGEAAGAVOELARALAHLLAEQERARAEAEQDQARVLAQLL 110
DQ 64 -----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAGAA 104
QY 111 RVWGAPRNSDPALGLDDDDPDAPAAQLARLRLARLDPAALAAQLVPAPVPAALRRPPV 170
DQ 105 GAWGG-----DREPPAA-----GPRALGP-----PAEPLLAANGTVPS 138
QY 171 YDGPAGPDAAEAGDETP 188
DQ 139 WPTAPV-PSAGEPGEAP 155

RESULT 7
US-09-795-686-3
; Sequence 3, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-686-3

Query Match      9.4%; Score 124.5; DB 10; Length 418;
Best Local Similarity 28.8%; Pred. No. 0.11;
Matches 57; Conservative 14; Mismatches 64; Indels 63; Gaps 9;

QY 2 AGSPILWGPAGVG-----LLVLLVLLGLFRPPPPALCARPVKEPRGLSAAAPPL 50
DQ 10 SGRP-----GPRARPGSAARSPPPLPLLLLLGLTAAAPGAA--AGNEAAPAGASVCSP- 63
QY 51 AETGAPRRFRSVPRGEAAGAVOELARALAHLLAEQERARAEAEQDQARVLAQLL 110
DQ 64 -----SVGSVQELAQRAAVVIEGKVHPQRR--QQGALDRKAAAGAA 104
QY 111 RVWGAPRNSDPALGLDDDDPDAPAAQLARLRLARLDPAALAAQLVPAPVPAALRRPPV 170
DQ 105 GAWGG-----DREPPAA-----GPRALGP-----PAEPLLAANGTVPS 138
QY 171 YDGPAGPDAAEAGDETP 188
DQ 139 WPTAPV-PSAGEPGEAP 155

RESULT 6
US-09-795-668-3
; Sequence 3, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
```



```
; SEQ ID NO 47
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-023-523-47

Query Match          9.0%; Score 118.5; DB 12; Length 550;
Best Local Similarity 28.4%; Pred. No. 0.39;
Matches 76; Conservative 14; Mismatches 103; Indels 75; Gaps 12;

QY 10 PRAGGVGLLVLLGLFRPPALCARPVKEPRG--LSAASP-----PLAETGAP 56
DQ 103 PRGAT-----PPA-----PPRPRGGPAAAPPTTAPPPPPAFAAANAAP 145
QY 57 RFRERSVPRGEAGAVOELARALAHLEAEQERARAEQAEEDQARVLAQLLRVWGAP 116
DQ 146 AR-----APRAAAAAA-----ATAPSPGPAQPGPRA-----QRAAPLAAPPPAPAAP 189
QY 117 RNSDPALGLDDDDPAAPAAQLAR-ALLRARLDPAALAAQLVPAPVPAALPRPPVYDDGP 175
DQ 190 PAAPAGPRPRAPPAANAARSPPLPPPPPPAPQOQQOQQPPPPPPPPQOQPP-----P 244
QY 176 AGPDAEAGDETPDVPDELLRLLGLRILAGSADSEG-----VAAPRRLRRAADHDV 226
DQ 245 EGGAARAGGPAPVSLREVYVILGGSGAGRLTRGRVQGLLEEAARGLERTR----- 300
QY 227 GSELPPEVGLALLVRKRLTETAPQVPA 254
DQ 301 -----LGA-LALPRGDRPGRAPPA 318

RESULT 11
US-09-825-288A-2
; Sequence 2, Application US/09825288A
; Publication No. US20020192822A1
; GENERAL INFORMATION:
; APPLICANT: ROIZMAN, BERNARD
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
; FILE REFERENCE: ARCD:317USC1
; CURRENT APPLICATION NUMBER: US/09/825,288A
; PRIOR FILING DATE: 2001-04-02
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 08/690,473
; PRIOR FILING DATE: 1996-07-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1298
; TYPE: PRT
; ORGANISM: HERPES VIRUS, TYPE 1
US-09-825-288A-2

Query Match          9.0%; Score 118.5; DB 9; Length 1298;
Best Local Similarity 24.9%; Pred. No. 1;
Matches 88; Conservative 24; Mismatches 121; Indels 121; Gaps 15;

QY 2 AGSP-LIWGPRAGGVGLLVLLGLFRPPALCARPVKEPR----- 41
DQ 355 SGAPAAVAPDELGDAAQOVALITRLTYTPDAEAMGWLONPRVVGVDVALDQACFRISGAA 414
QY 42 -----GLSAASPLAETGAPRFRRSVPRGEAGAVQ-----ELARAL 79
DQ 415 RNSSSFTTGSVARVPHLGMAAGRGWGLAHAAAVAMRRYDRAQKGFLLTSLRRAY 474
QY 80 AHLEAER-----QERARAPAEQOQ-----ARVLAQLLRVWGAP 116
DQ 475 APLLARENAALTGAAGSPGAGADDEGVAAVAAAPGERAVPAGYGAAGIILALGRLSAAP 534
QY 117 RNSDPALGLDDDDPAP-----AAQLAR-----ALLRARLDPAALAAQLVPAPV 159
DQ 535 --ASPAGG--DDPDAARHADDDAGRRAGRAVAVECLAACRGILEALEAGFGDGLAAV 590

; SEQ ID NO 160
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-077-130-2

Query Match          8.7%; Score 114.5; DB 10; Length 492;
Best Local Similarity 25.1%; Pred. No. 0.66;
Matches 78; Conservative 19; Mismatches 87; Indels 127; Gaps 16;

QY 160 PA-AALRPRPPVYDDGPGAGPDAEAGDETPDVPDELLRLLGRI----- 202
DQ 591 PGLAGARPASPAPRPEGPAGP-----ASPPPHADAPRLRAWLRELFRVDRDALVLMRLRGDL 646
QY 203 -LAGSADSEGVAAAPRRLRRAADHDVSELPEGVIG-ALLRVKRLTETAPQVPA 254
DQ 647 RVAGGSEA-AVAARAVSLVA-----GALGPALPRDPRLPSSAAAAA 688

RESULT 12
US-09-764-898-206
; Sequence 206, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - CONSULT PALM OR FILE WRAPPER
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-898-206

Query Match          8.7%; Score 114.5; DB 10; Length 492;
Best Local Similarity 25.1%; Pred. No. 0.66;
Matches 78; Conservative 19; Mismatches 87; Indels 127; Gaps 16;

QY 28 PPPALCARPVKEPRGLSASP-----PLAETGAPRFR-----RS 62
DQ 194 PPEDAPAP-----PTPGPAASPEQLSFRERKYEFELEVRVPOAE-GPKRVSLVGCADDLRK 249
QY 63 VPGEAAGAVOELARALAHLEAEQERARAE-----QEAEDQARVLAQLLRVWGAPR 117
DQ 250 MQEEARKLQKQRAQMLREAAEAGAEARLALDGTGLCEEQEDEQP-----PWASPS 301
QY 118 -----NSDPALG-----LDDDDPAALAAQLARALLRA----- 143
DQ 302 PTKSQSPASPPPLGGGAPVTKAERRHOERLRVQSPPEPPAPERALSPEALRALEAEKRA 361
QY 144 -----RLDPAALAAQLV-----PAPVPAALRPRP-PVYDDG 174
DQ 362 LWRAARMKSLQDALRAQWVLSRQSGRGTRGPLERLAEAPSPAPT-----PSTPTVEDLG 417
QY 175 P-----AGPD-AEEAGDETPDVP-----ELLRYLLGLRILAGSADSEGVAA-APR 216
DQ 418 PQTSTSPGRSLSPDFAELRSLEFSPSPGQEDGEVALVLLGRSPGAVGPDVALCSSLR 477
QY 217 RLRRADHDVG 227
DQ 478 RPVRPGRRGGLG 488

RESULT 13
US-10-077-130-2
; Sequence 2, Application US/10077130
; Patent No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE REFERENCE: MPI2001-047P1RCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
```

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 2630

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-077-130-2

Query Match 8.5%; Score 112; DB 9; Length 2630;

Best Local Similarity 27.7%; Pred. No. 6.4;

Matches 81; Conservative 23; Mismatches 98; Indels 90; Gaps 15;

QY 28 PPPALCARP-VKEPRGLSASPLP---AETGAPRRFRSVPRGEAAGAV-----72

DB 1804 PGPSLDAEGWTQEAEDLSDTPTLQRPQEQATMRKFSLG-GRGGYAGVAGYGTFAFGGDA 1862

QY 73 -----QELARALAHLEAEQERARAEQAEDQQAARVLAQLLRVWG-----114

DB 1863 GGMGLGGPMMARIAVAVSQ-SEEEQEAEARAESEEQEAEARAESEPLQVVSARPPVEYGR 1921

QY 115 APRNSDPA-----LG-----LDDDDPAPAAQLARALLRLDPAAL-----150

DB 1922 APTRSSPEPTWEDIGQVSLVQIRDLSDGDAEAAADTI-SLIDISEVDPAYLNLSLDYDIKYL 1980

QY 151 -----AAQLVPAPVPA-----ALRPPVPYDDGPGAGDAE-EAGDETDPVDPE 193

DB 1981 PFEWIFRKVPKSAQPEP-PSMAEEELAEPEPTWMPWPGELGPHAGLEITESESDVDAL 2039

QY 194 LLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSGLPPE--GVLGALLRVK 243

DB 2040 LAEAAVGRKRKWSPPS-----RSLFHPGGRHLPLDEPAELGLRERVK 2081

RESULT 14

US-10-077-130-5

; Sequence 5, Application US/100771130

; Patent No. US20020168742A1

; GENERAL INFORMATION:

; APPLICANT: Kapeller-Libermann, Rosana

; APPLICANT: Acton, Susan L.

; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family

; TITLE OF INVENTION: Members and Uses Thereof

; FILE REFERENCE: MPI2001-047P1RCPl(M)

; CURRENT APPLICATION NUMBER: US/10/077,130

; PRIOR FILING DATE: 2002-02-15

; PRIOR FILING DATE: 2001-02-15

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 7968

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-077-130-5

Query Match 8.5%; Score 112; DB 9; Length 7968;

Best Local Similarity 27.7%; Pred. No. 22;

Matches 81; Conservative 23; Mismatches 98; Indels 90; Gaps 15;

QY 28 PPPALCARP-VKEPRGLSASPLP---AETGAPRRFRSVPRGEAAGAV-----72

DB 7142 PGPSLDAEGWTQEAEDLSDTPTLQRPQEQATMRKFSLG-GRGGYAGVAGYGTFAFGGDA 7200

QY 73 -----QELARALAHLEAEQERARAEQAEDQQAARVLAQLLRVWG-----114

DB 7201 GGMGLGGPMMARIAVAVSQ-SEEEQEAEARAESEEQEAEARAESEPLQVVSARPPVEYGR 7259

QY 115 APRNSDPA-----LG-----LDDDDPAPAAQLARALLRLDPAAL-----150

DB 7260 APTRSSPEPTWEDIGQVSLVQIRDLSDGDAEAAADTI-SLIDISEVDPAYLNLSLDYDIKYL 7318

QY 151 -----AAQLVPAPVPA-----ALRPPVPYDDGPGAGDAE-EAGDETDPVDPE 193

DB 7319 PFEWIFRKVPKSAQPEP-PSMAEEELAEPEPTWMPWPGELGPHAGLEITESESDVDAL 7377

QY 194 LLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSGLPPE--GVLGALLRVK 243

DB 7378 LAEAAVGRKRKWSPPS-----RSLFHPGGRHLPLDEPAELGLRERVK 7419

RESULT 15

US-09-976-740-43

; Sequence 43, Application US/09976740

; Publication No. US20020194633A1

; GENERAL INFORMATION:

; APPLICANT: Lees, Ann M.

; APPLICANT: Lees, Robert S.

; APPLICANT: Law, Simon W.

; APPLICANT: Arjona, Anibal A.

; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING

; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING

; TITLE OF INVENTION: ATHEROSCLEROSIS

; FILE REFERENCE: 10797-004001

; CURRENT APPLICATION NUMBER: US/09/976,740

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: 09/616,289

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: US 08/979,608

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: US 60/031,930

; PRIOR FILING DATE: 1996-11-27

; PRIOR APPLICATION NUMBER: US 60/048,547

; NUMBER OF SEQ ID NOS: 53

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 43

; LENGTH: 538

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-976-740-43

Query Match 8.4%; Score 110.5; DB 9; Length 538;

Best Local Similarity 26.7%; Pred. No. 1.4;

Matches 77; Conservative 17; Mismatches 99; Indels 95; Gaps 16;

QY 1 MAGSLLMGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSASPP-----LAETGA 55

DB 1 MAGPPAL-----PPPETAAAAATTAASSSSAASPHYQEWILDTID 40

QY 56 PRFRFRSVPRGENAGAVOELARALAHLEAEQERARAEQAEDQQAARVLAQLLRV--- 112

DB 41 SLRSRKARP-----DLERICR-WVRRRHGPEPEPTRAELEKLIQRA-----VLRSYK 88

QY 113 -----WGAPRNSDPALGLDDDDPDAP--AAQALARALLRLDPAALAAQLVPAPVPA--A 163

DB 89 GSISYRNAARVQPPRRGA--TPPAPPAPRGAPAAAAAAPPPTAPPAPVPAVAAAAA 146

QY 164 LRPR-----PPVYDDGPA--GPDAAEAGD-ETPDVDPELLRYLLGRILAGSADSEGV 212

DB 147 RAPRAAAAAATAPP--SPGAPQGPRAQRAAPLAAPPAP-----AAPPAV 190

QY 213 AAPRLRRAADHDVGSGLPPEGVLGALLRVKRLTETPAQVPPARRLLPP 260

DB 191 APPAGPRRAPPAVAAREPP-----LPPPPQPPA-----PP 221

Search completed: April 4, 2003, 14:26:42

Job time : 42 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:19:00 ; Search time 74 Seconds
(without alignments)
468.178 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 260
Sequence: 1 MAGSPLLMGPRAGGVLVLL.....RVKRLTTPAPQVPARRLLPP 260

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 10

Total number of hits satisfying chosen parameters: 14

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: *

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: *

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: *

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: *

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: *

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: *

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: *

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: *

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: *

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: *

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: *

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: *

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: *

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: *

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: *

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: *

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: *

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: *

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: *

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: *

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: *

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: *

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	260	100.0	260	20	AAW84597
2	260	100.0	260	21	AAW71959
3	260	100.0	260	22	AAW39005
4	260	100.0	260	22	AAW88596
5	200	76.9	223	22	AAW39006
6	51	19.6	148	22	AAW06088
7	51	19.6	148	23	ABG33910
8	51	19.6	149	21	AAW87111
9	51	19.6	166	21	AAW87199
10	51	19.6	166	22	AAW06176

11	51	19.6	166	23	ABG34000
12	33	12.7	258	21	AAW71960
13	33	12.7	262	21	AAW71962
14	12	4.6	69	21	AAW71961

ALIGNMENTS

RESULT 1
AAW84597

ID AAW84597 standard; Protein; 260 AA.

XX AAW84597;

XX 03-JUN-1999 (first entry)

DT Amino acid sequence of the human Tango-81 protein.

DE Human; Tango-81; host cell; recombinant protein; antibody;

KW receptor; specific binding agent; probe; primer; hybridisation;

KW amplification; mutation; genetic mapping.

XX Homo sapiens.

XX WO9906427-A1.

XX 11-FEB-1999.

XX 04-AUG-1998; 98WO-US16241.

XX 04-AUG-1997; 97US-0054645.

XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

XX McCarthy SA;

XX WPI; 1999-153693/13.

XX N-PSDB; AAV68668.

XX New nucleic acid encoding human Tango-78, -79 and -81 proteins -

XX useful for diagnosis and treatment of Tango-associated diseases

XX Claim 1; Fig 3; 67pp; English.

XX This is the amino acid sequence of the human Tango-81 protein

XX used in the method of the invention. Host cells containing

XX the Tango protein are used to produce recombinant proteins for

XX raising antibodies. It is also used in identifying specific

XX binding agents (including cognate receptors), which can be used to

XX determine amounts of recombinant protein in cells or

XX therapeutically. Antibodies or other specific binding agents, are

XX used to detect recombinant proteins and fragments of the Tango

XX nucleic acid sequence are also used for genetic mapping and

XX chromosome identification, and as antisense, ribozyme or

XX triplex-forming therapeutics. Antibodies may also be used to generate

XX anti-idiotypic antibodies.

XX Sequence 260 AA;

XX Query Match 100.0%; Score 260; DB 20; Length 260;

XX Best Local Similarity 100.0%; Pred. No. 7.8e-213;

XX Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLMGPRAGGVLVLLLLGLFRPPALCARPKVPEKGLSAASPPLAETGAPRRR 60

DB 1 MAGSPLLMGPRAGGVLVLLLLGLFRPPALCARPKVPEKGLSAASPPLAETGAPRRR 60

QY 61 RSVPRGEAGAVQELARALAHLEAERQERARAEAEQAEQARVLAQLLRVWGAPRNSD 120
Db 61 RSVPRGEAGAVQELARALAHLEAERQERARAEAEQAEQARVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPDAPAAQARALLRARLDPAALAAQLVPAPVPAALRPVPPVDDGAPGPA 180
Db 121 PALGLDDDDPDAPAAQARALLRARLDPAALAAQLVPAPVPAALRPVPPVDDGAPGPA 180
QY 181 EAGDETPDVDPPELLRYLLGRILAGSADSEGVAAAPRRLRRADHDVGSSELPEGVLGALL 240
Db 181 EAGDETPDVDPPELLRYLLGRILAGSADSEGVAAAPRRLRRADHDVGSSELPEGVLGALL 240
QY 241 RVKRLTPAPQVPAARRLLPP 260
Db 241 RVKRLTPAPQVPAARRLLPP 260
RESULT 2
ID AAY71959 standard; Protein; 260 AA.
AC AAY71959;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human pituitary hormone, pituitrone.
XX
KW Human; pituitary; pituitrone; therapy; immune disorder; anaemia;
KW George syndrome; haematopoietic cell; ataxia telangiectasia;
KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW SLE; hyperproliferative disorder; gene therapy; neoplasm;
KW infectious disease; immunomodulatory; cytostatic; antimicrobial.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..33
FT Protein 34..260
FT Label= Mature_human_pituitrone
FT Region 36..44
FT /note= "Highly immunogenic"
FT Region 55..68
FT /note= "Highly immunogenic"
FT Region 116..121
FT /note= "Highly immunogenic"
FT Region 123..131
FT /note= "Highly immunogenic"
FT Region 164..182
FT /note= "Highly immunogenic"
FT Region 185..193
FT /note= "Highly immunogenic"
FT Region 206..211
FT /note= "Highly immunogenic"
FT Region 225..233
FT /note= "Highly immunogenic"
FT Region 253..258
FT /note= "Highly immunogenic"
XX
PN WO200066778-A1.
XX
PD 09-NOV-2000.
XX
PF 27-APR-2000; 2000WO-US11211.
XX
PR 30-APR-1999; 99US-0131966.
XX
FA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, N1 J;
XX
XX WPI; 2000-687547/67.
DR

DR N-PSDB; AAD02075.
XX
PT Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
PT for diagnosing, preventing and treating e.g. immune disorders,
PT hyperproliferative disorders and blood coagulation disorders -
XX
PS Claim 11; Fig 1; 277pp; English.
XX
CC The present sequence is a human pituitary hormone, pituitrone (clone
CC HKGDL36). Pituitrone is highly expressed in pituitary
CC gland and is also expressed in brain tissues, spinal cord and kidney.
CC It may be used as antigens in the production of antibodies against
CC pituitrone and in assays to identify modulators.. Pituitrone CDNAs are
CC also useful in gene therapy. Pituitrone may be used in the prevention,
CC treatment and diagnosis of diseases associated with inappropriate
CC pituitrone expression. It may be useful in treating disorders related
CC to reproductive and renal system, immune disorders, disorders of
CC haematopoietic cells (e.g., anaemia, Digorge syndrome, ataxia
CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation
CC disorders, autoimmune disorders (e.g. Addison's disease, multiple
CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
CC diseases.
XX
SQ Sequence 260 AA;
Query Match 100.0%; Score 260; DB 21; Length 260;
Best Local Similarity 100.0%; Pred. No. 7.8e-213;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAGSPLLMGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPLAETGAPRRFR 60
Db 1 MAGSPLLMGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPLAETGAPRRFR 60
Qy 61 RSVPRGEAGAVQELARALAHLEAERQERARAEAEQAEQARVLAQLLRVWGAPRNSD 120
Db 61 RSVPRGEAGAVQELARALAHLEAERQERARAEAEQAEQARVLAQLLRVWGAPRNSD 120
Qy 121 PALGLDDDDPDAPAAQARALLRARLDPAALAAQLVPAPVPAALRPVPPVDDGAPGPA 180
Db 121 PALGLDDDDPDAPAAQARALLRARLDPAALAAQLVPAPVPAALRPVPPVDDGAPGPA 180
Qy 181 EAGDETPDVDPPELLRYLLGRILAGSADSEGVAAAPRRLRRADHDVGSSELPEGVLGALL 240
Db 181 EAGDETPDVDPPELLRYLLGRILAGSADSEGVAAAPRRLRRADHDVGSSELPEGVLGALL 240
Qy 241 RVKRLTPAPQVPAARRLLPP 260
Db 241 RVKRLTPAPQVPAARRLLPP 260
RESULT 3
AAM39005
ID AAM39005 standard; Protein; 260 AA.
XX
AC AAM39005;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2150.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.

```

XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0523217.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR N-PSDB; AA158161.
XX XX
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Example 4; SEQ ID NO 2150; 10078pp; English.
XX XX
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AA38642-AA42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX XX
XX SQ Sequence 260 AA;
Query Match 100.0%; Score 260; DB 22; Length 260;
Best Local Similarity 100.0%; Pred. No. 7.8e-213;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGSPLLWGPRAGGVGLLVLLGLFRPPPPALCARVPKPRGLSAASPPPLAETGAPRRFR 60
Db 1 MAGSPLLWGPRAGGVGLLVLLGLFRPPPPALCARVPKPRGLSAASPPPLAETGAPRRFR 60
QY 61 RSVPRGEAGAGVQELARALAHLEAEERQERARAEAEQEQARVLAQLLRVWGAPRNSD 120
Db 61 RSVPRGEAGAGVQELARALAHLEAEERQERARAEAEQEQARVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPAPAAQALARALLRLRLDPAALAAQLVPAPVPAALRRPPVYDDGPAGPDA 180
Db 121 PALGLDDDDPAPAAQALARALLRLRLDPAALAAQLVPAPVPAALRRPPVYDDGPAGPDA 180
QY 181 EAGDETPDVPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPPGVLGALL 240
Db 181 EAGDETPDVPDPELLRYLLGRILAGSADSEGVAAAPRRRLRAADHDVGSLEPPGVLGALL 240
QY 241 RVKRLETPAPQVPAARLLPP 260
Db 241 RVKRLETPAPQVPAARLLPP 260
RESULT 4
AAB88596
ID AAB88596 standard; Protein; 260 AA.

```

```

XX AC AAB88596;
XX DT 04-JUN-2001 (first entry)
XX DE Human hydrophobic domain containing protein clone HP10756 #100.
XX KW Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;
XX KW antianemic; vulnery; antiulcer; osteopathic; anti-inflammatory;
XX KW cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;
XX KW HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;
XX KW inflammatory bowel disease; nutritional supplement; appetite; vaccine;
XX KW behavioural characteristic; immune response.
XX OS Homo sapiens.
XX PN WO200112660-A2.
XX PD 22-FEB-2001.
XX PF 10-AUG-2000; 2000WO-JP05356.
XX PR 17-AUG-1999; 99JP-0230344.
XX PR 07-SEP-1999; 99JP-0252551.
XX PR 01-OCT-1999; 99JP-0281132.
XX PR 22-OCT-1999; 99JP-0301624.
XX PR 04-NOV-1999; 99JP-0313877.
XX PA (SAGA ) SAGAMI CHEM RES CENT.
XX PA (PROT-) PROTEGENE INC.
XX PI Kato S, Kimura T;
XX DR WPI: 2001-160059/16.
XX DR N-PSDB; AAF94486.
XX XX
XX PT Human proteins with hydrophobic domains and the DNAs which encode them
XX PT are useful for treating autoimmune disorders, burns and tumors and for
XX PT screening novel pharmaceuticals -
XX PS Claim 1; Page 422-423; 518pp; English.
XX XX
XX CC AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
XX CC AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
XX CC anti-HIV, neuroprotective, antianemic, vulnery, antiulcer,
XX CC osteopathic, anti-inflammatory and cytostatic activities, and can be
XX CC used in gene therapy. (I) can be used as pharmaceuticals and as antigens
XX CC to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
XX CC probes for genetic diagnosis and gene sources for gene therapy or for
XX CC producing (I) in large quantities. Cells containing (II) are used for
XX CC the detection of ligands or receptors corresponding to membrane or
XX CC secretory proteins and to screen small molecule novel pharmaceuticals.
XX CC Antibodies directed to (I) can be used for the detection, quantification
XX CC and purification of (I). Activities of (I) may include cytokine and cell
XX CC proliferation/differentiation function, immune stimulating or suppressing
XX CC activity, haematopoiesis regulating activity, tissue growth activity,
XX CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory
XX CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.
XX CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
XX CC inflammatory bowel disease and tumors. (I) and (II) can also be used for
XX CC wound healing, as nutritional sources or supplements e.g. as amino acid,
XX CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
XX CC processing and utilisation of dietary fat, protein, carbohydrate,
XX CC vitamins and minerals, to effect behavioural characteristics, to affect
XX CC appetite, and can act as antigens in vaccines to raise an immune response
XX CC to the protein or another material cross-reactive with the protein.
XX SQ Sequence 260 AA;
Query Match 100.0%; Score 260; DB 22; Length 260;
Best Local Similarity 100.0%; Pred. No. 7.8e-213;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



```
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
XX WPI; 2001-425865/45.
DR N-PSDB; AAD11677.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
PT
XX Claim 11; Page 773-774; 864pp; English.
XX
XX AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted
CC protein genes, and AAE06041-AAE06132 represent the proteins they encode.
CC AAE06133-AAE06205 represent human secreted protein fragments.
CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 71 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiosenic disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein of
CC the invention.
XX
XX Sequence 148 AA;
SQ
Query Match. 19.6%; Score 51; DB 22; Length 148;
Best Local Similarity 100.0%; Pred. No. 2e-35;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGSPLLWGPAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
DB 1 MAGSPLLWGPAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
|||||
RESULT 7
ABG33910
ID ABG33910 standard; Protein; 148 AA.
XX
XX ABG33910;
XX
XX 15-JUL-2002 (first entry)
DT
XX Human secreted protein encode by gene 48.
DE
XX Human: secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antrineumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; opthalmalagical; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
```

```
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
XX Homo sapiens.
XX WO200226931-A2.
XX
XX 04-APR-2002.
XX
XX 24-SEP-2001; 2001WO-US29871.
XX
XX 25-SEP-2000; 2000US-234925P.
PR 12-JAN-2001; 2001WO-US00911.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
XX WPI; 2002-362489/39.
DR N-PSDB; ABK69773.
XX
XX Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT the polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT asthma
XX
XX Claim 11; Page 1252-1253; 1478pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC for example autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. Many other diseases and
CC disorders are listed in the specification. The polypeptides can also be
CC used to aid wound healing an epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence represents a novel human secreted protein of the invention.
XX
XX Sequence 148 AA;
SQ
Query Match 19.6%; Score 51; DB 23; Length 148;
Best Local Similarity 100.0%; Pred. No. 2e-35;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGSPLLWGPAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
DB 1 MAGSPLLWGPAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
|||||
RESULT 8
AY87111
ID AAY87111 standard; Protein; 149 AA.
XX
XX AAY87111;
XX/ 09-MAY-2000 (first entry)
DT
XX
```

Human secreted protein sequence SEQ ID NO:150.

DE XX Human; secreted protein; diagnosis; cytostatic; cytostatic; immunosuppressive;
KW antinflammatory; nontropic; neuroprotective; antiallergic; cancer;
KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
KW foetal deficiency; blood disorder; immune system disorder; arthritis;
KW autoimmune disease; hepatic disease; renal disease; inflammation;
KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
KW metabolic disorder; food additive; preservative.

XX Homo sapiens.

OS WO200004140-A1.

XX 27-JAN-2000.

XX 14-JUL-1999; 99WO-US15849.

XX 15-JUL-1998; 98US-0092921.

XX 15-JUL-1998; 98US-0092922.

XX 15-JUL-1998; 98US-0092956.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
PI Mucenski M, Endress GA, Soppet DR;

XX WPI: 2000-161128/14.

DR N-PSDB; AAZ98064.

PT New isolated human genes, useful for diagnosis and treatment of, e.g.
PT cancers, neurological or blood disorders

XX Claim 11; Page 406-407; 494pp; English.

XX The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the
CC human secreted proteins given in AAZ97064 to AAZ97223. Human secreted
CC protein can have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic;
CC immunosuppressive; antinflammatory; nontropic; neuroprotective; and
CC antiallergic. The polynucleotides and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Human secreted protein s and their polynucleotides can
CC be used for developing products for the diagnosis or treatment of cancer,
CC tumours, neurodegenerative disorders, developmental abnormalities and
CC foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The
CC proteins or polynucleotides can also be used as food additives or
CC preservatives. The proteins are also useful for identifying their
CC binding partners. AAZ98008 to AAZ98016 and AAZ97063 are sequence used in
CC the exemplification of the present invention.

XX Sequence 149 AA;

Query Match 19.6%; Score 51; DB 21; Length 149;
Best Local Similarity 100.0%; Pred. No. 2e-35;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLWGPAGGVLVLLLLGLFPPPPALCARPVKEPRGLSAAAGPPLA 51
|||||
DB 1 MAGSPLLWGPAGGVLVLLLLGLFPPPPALCARPVKEPRGLSAAAGPPLA 51

RESULT 9
AAZ98199
ID AAZ98199 standard; Protein; 166 AA.
XX
AC AAZ98199;
XX
DT 09-MAY-2000 (first entry)
XX
DE Human secreted protein sequence SEQ ID NO:238.
XX
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antinflammatory; nontropic; neuroprotective; antiallergic; cancer;
KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
KW foetal deficiency; blood disorder; immune system disorder; arthritis;
KW autoimmune disease; hepatic disease; renal disease; inflammation;
KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
KW metabolic disorder; food additive; preservative.

OS Homo sapiens.

XX WO200004140-A1.

XX 27-JAN-2000.

XX 14-JUL-1999; 99WO-US15849.

XX 15-JUL-1998; 98US-0092921.

XX 15-JUL-1998; 98US-0092922.

XX 15-JUL-1998; 98US-0092956.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
PI Mucenski M, Endress GA, Soppet DR;

XX WPI: 2000-161128/14.

DR N-PSDB; AAZ98064.

PT New isolated human genes, useful for diagnosis and treatment of, e.g.
PT cancers, neurological or blood disorders

XX Claim 11; Page 406-407; 494pp; English.

XX The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the
CC human secreted proteins given in AAZ97064 to AAZ97223. Human secreted
CC protein can have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic;
CC immunosuppressive; antinflammatory; nontropic; neuroprotective; and
CC antiallergic. The polynucleotides and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Human secreted protein s and their polynucleotides can
CC be used for developing products for the diagnosis or treatment of cancer,
CC tumours, neurodegenerative disorders, developmental abnormalities and
CC foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The
CC proteins or polynucleotides can also be used as food additives or
CC preservatives. The proteins are also useful for identifying their
CC binding partners. AAZ98008 to AAZ98016 and AAZ97063 are sequence used in
CC the exemplification of the present invention.

XX Sequence 149 AA;

Query Match 19.6%; Score 51; DB 21; Length 149;
Best Local Similarity 100.0%; Pred. No. 2e-35;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLWGPAGGVLVLLLLGLFPPPPALCARPVKEPRGLSAAAGPPLA 51
|||||
DB 1 MAGSPLLWGPAGGVLVLLLLGLFPPPPALCARPVKEPRGLSAAAGPPLA 51

XX Sequence 166 AA;
SQ Query Match 19.6%; Score 51; DB 21; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.2e-35;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAGSPLLWGPAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
|||||
Db 19 MAGSPLLWGPAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 69

RESULT 10
AAE06176
ID AAE06176 standard; Protein; 166 AA.
AC AAE06176;
DT 24-SEP-2001 (first entry)
XX Human gene 48 encoded secreted protein fragment, SEQ ID NO:238.
DE Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angioinetic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnerability; binding partner identification;
KW gene therapy.
XX Homo sapiens.
OS
XX WO200151504-A1.
PN
XX 19-JUL-2001.
PD
XX 12-JAN-2001; 2001WO-US00911.
PF
XX 13-JAN-2000; 2000US-0482273.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX WPI; 2001-425865/45.
DR
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
PS Disclosure; Page 122; 864pp; English.
XX

AD11630-AD11721 represent cDNAs corresponding to 71 human secreted
CC protein genes, and AAE06041-AAE06132 represent the proteins they encode.
CC AAE06133-AAE06205 represent human secreted protein fragments.
CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 71 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angioinetic disorders, kidney disorders, gastrointestinal disorders,

CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein fragment
CC referred to in the disclosure of the invention.
XX
XX Sequence 166 AA;
SQ Query Match 19.6%; Score 51; DB 22; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.2e-35;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAGSPLLWGPAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 51
|||||
Db 19 MAGSPLLWGPAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPPLA 69

RESULT 11
ABG34000
ID ABG34000 standard; Protein; 166 AA.
XX
AC ABG34000;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein #2 encoded by gene 48.
XX
DE Human; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; opthalmological; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
XX Homo sapiens.
OS
XX WO200226931-A2.
PN
XX 04-APR-2002.
PD
XX 24-SEP-2001; 2001WO-US29871.
PF
XX 25-SEP-2000; 2000US-234925P.
PR
XX 12-JAN-2001; 2001WO-US00911.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX WPI; 2002-362489/39.
DR
XX N-PSDB; ABK69773.
DR
XX Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT the polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT asthma
XX
XX Disclosure; Page 1316-1317; 1478pp; English.
PS
XX The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human

secreted protein (and its fragment, domain, epitope, variant, secreted form and species variant). Also included are a recombinant vector comprising the nucleic acid, a recombinant host cell comprising the vector, an antibody against the secreted protein, a recombinant host cell that expresses the secreted protein and a method of identifying a binding partner of the secreted protein. The nucleic acid and protein are used to prevent, diagnose, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. Many other diseases and disorders are listed in the specification. The polypeptides can also be used to aid wound healing an epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The present sequence represents a novel human secreted protein of the invention.

Sequence 166 AA;

Query Match 19.6%; Score 51; DB 23; Length 166;

Best Local Similarity 100.0%; Pred. No. 2.2e-35;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLWGPAGGVLLVLLGLFRPPPCALCARVKEPRGLSAASPPLA 51

|||||

DB 19 MAGSPLLWGPAGGVLLVLLGLFRPPPCALCARVKEPRGLSAASPPLA 69

RESULT 12

AAAY71960

ID AAY71960 standard; Protein; 258 AA.

XX AC AAY71960;

XX DT 26-MAR-2001 (first entry)

XX Mouse pituitary hormone, pituitrone.

XX Mouse; pituitary; pituitrone; therapy; immune disorder; anaemia; Digorge syndrome; haematopoietic cell; ataxia telangiectasia; Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease; autoimmune disorder; multiple sclerosis; systemic lupus erythematosus; SLE; hyperproliferative disorder; gene therapy; neoplasm; infectious disease; immunomodulatory; cytostatic; antimicrobial.

XX OS Mus sp.

XX Key Location/Qualifiers

XX Peptide 1..34

FT /label= Signal_peptide

FT /note= "An alternative signal sequence referred in

FT page 11"

FT Peptide 1..33

FT /label= Signal_peptide

FT Protein 35..258

FT /note= "Mature mouse pituitrone"

FT /note= "An alternative mature protein referred in

FT page 11"

FT 34..258

FT /note= "Mature mouse pituitrone"

XX WO2000066778-A1.

XX PD 09-NOV-2000.

XX 27-APR-2000; 2000WO-US11211.

XX 30-APR-1999; 99US-0131966.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, N1 J;

XX WPI; 2000-687547/67.

XX N-PSDB; AAD02076.

XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful for diagnosing, preventing and treating e.g. immune disorders, hyperproliferative disorders and blood coagulation disorders -

XX Claim 11; Fig 2; 277pp; English.

XX The present sequence is a mouse pituitary hormone, pituitrone. Pituitrone is highly expressed in pituitary gland and is also expressed in brain tissues, spinal cord and kidney. It may be used as antigens in the production of antibodies against pituitrone and in assays to identify modulators. Pituitrone cDNAs are also useful in gene therapy. Pituitrone may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate pituitrone expression. It may be useful in treating disorders related to reproductive and renal system, immune disorders, disorders of haematopoietic cells (e.g., anaemia, Digorge syndrome, ataxia telangiectasia and Wiskott-Aldrich disorder), blood coagulation disorders, autoimmune disorders (e.g. Addison's disease, multiple sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative disorders (e.g. neoplasms of bone, liver and pancreas) and infectious diseases.

XX Sequence 258 AA;

Query Match 12.7%; Score 33; DB 21; Length 258;

Best Local Similarity 100.0%; Pred. No. 6.5e-20;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 DDDPAPAAQLARALLRRLRDLPAALAAQLVPAP 158

|||||

DB 126 DDDPAPAAQLARALLRRLRDLPAALAAQLVPAP 158

RESULT 13

AAAY71962

ID AAY71962 standard; Protein; 262 AA.

XX AC AAY71962;

XX DT 26-MAR-2001 (first entry)

XX Consensus sequence of human, mouse and rat pituitrone.

XX Rat; human; mouse; pituitary; pituitrone; therapy; immune disorder; Digorge syndrome; haematopoietic cell; ataxia telangiectasia; Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease; autoimmune disorder; multiple sclerosis; systemic lupus erythematosus; SLE; hyperproliferative disorder; gene therapy; neoplasm; anaemia; infectious disease; immunomodulatory; cytostatic; antimicrobial.

XX Rattus sp.

XX OS Mus sp.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 8..12

FT /label= Unknown

FT Misc-difference 17

FT /label= Unknown

FT Misc-difference 26

FT /label= Unknown

FT Misc-difference 28

FT /label= Unknown

FT Misc-difference 31

FT /label= Unknown


```

XX 30-APR-1999; 99US-0131966.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX PI Ruben SM, N1 J;
XX DR WPI; 2000-687547/67.
XX DR N-PSDB; AAD02077.
XX
XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
PT for diagnosing, preventing and treating e.g. immune disorders,
PT hyperproliferative disorders and blood coagulation disorders -
XX
XX Claim 11: Fig 2; 277pp; English.
XX
XX The present sequence is a rat pituitary hormone, pituitrone.
CC Pituitrone is highly expressed in pituitary gland and is
CC also expressed in brain tissues, spinal cord and kidney. It may be
CC used as antigens in the production of antibodies against pituitrone
CC and in assays to identify modulators. Pituitrone cDNAs are
CC also useful in gene therapy. Pituitrone may be used in the prevention,
CC treatment and diagnosis of diseases associated with inappropriate
CC pituitrone expression. It may be useful in treating disorders related
CC to reproductive and renal system, immune disorders, disorders of
CC haematopoietic cells (e.g., anaemia, Digorge syndrome, ataxia
CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation
CC disorders, autoimmune disorders (e.g. Addison's disease, multiple
CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
CC diseases.
XX
SQ Sequence 69 AA;
Query Match 4.6%; Score 12; DB 21; Length 69;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 235 VLGALLRVKRL 246
Db 44 VLGALLRVKRL 55

```

Search completed: April 4, 2003, 14:28:29
Job time : 74 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 15:48:41 ; Search time 275 seconds
(without alignments)
8017.114 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcgacagagccagc.....gaggatccctaccctctggc 979

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	977	99.8	979	20 AAV68668	Nucleotide sequenc
2	961.4	98.2	1030	22 AAI58161	Human polynucleoti
3	961.4	98.2	1041	22 AAI59947	Human polynucleoti
4	961.4	98.2	1041	22 AAI59948	Human polynucleoti
5	961.4	98.2	1043	21 RAD02075	cDNA encoding huma
6	959.8	98.0	998	22 AAF94496	Human hydrophobic
7	912.8	93.2	1050	21 AAZ98064	Human secreted pro
8	912.8	93.2	1050	22 AAD11677	Human secreted pro
9	912.8	93.2	1050	24 ABK69773	Human secreted pro

10	776.8	79.3	780	22	AAF94486	Human hydrophobic
11	729.4	74.5	919	22	AAI58162	Human polynucleoti
12	565.4	57.8	991	21	AAD02076	cDNA encoding mous
13	474.2	48.4	4503	24	ABK69876	Human secreted pro
14	301.6	30.8	652	24	ABQ31154	Oligonucleotide fo
15	301.6	30.8	652	24	ABQ31155	Oligonucleotide fo
16	291.8	29.8	652	24	ABQ31156	Oligonucleotide fo
17	291.8	29.8	652	24	ABQ31157	Oligonucleotide fo
18	276.6	28.3	301	24	ABN95203	Gene #1701 used to
19	276.6	28.3	301	24	ABL68387	Kidney cancer rela
20	153.2	15.6	396	21	AAD02077	cDNA encoding rat
21	99.6	10.2	114955	20	AAX53491	Human adenosine A1
22	98.8	10.1	739	24	ABQ31160	Oligonucleotide fo
23	98.8	10.1	739	24	ABQ31161	Oligonucleotide fo
24	95.2	9.7	114955	20	AAX53491	Human adenosine A1
25	90.6	9.3	739	24	ABQ31158	Oligonucleotide fo
26	90.6	9.3	739	24	ABQ31159	Oligonucleotide fo
27	77.8	7.9	4403765	22	AAI99683	Mycobacterium tube
28	75.2	7.7	1614	22	AAH26499	Human low density
29	75.2	7.7	12425	22	AAH26499	Human low density
30	73.4	7.5	2561	22	AAH26500	Rabbit low density
31	71.6	7.3	4403765	22	AAI99683	Mycobacterium tube
32	70.4	7.2	3957	22	AAA09686	HSV-2 immediate ea
33	70.4	7.2	154746	24	AAD25519	Human herpesvirus
34	70.4	7.2	154746	24	AAD25519	Human herpesvirus
35	69.6	7.1	888	17	AAT59269	Streptomyces prist
36	69.6	7.1	2888	17	AAT59268	Streptomyces prist
37	69.4	7.1	2307	24	ABK32842	DNA encoding human
38	69.4	7.1	3957	22	AAA09686	HSV-2 immediate ea
39	69.4	7.1	4411529	22	AAI99682	Mycobacterium tube
40	68.8	7.0	125401	22	AAI17186	Streptomyces nous
41	68.6	7.0	1127	21	AAA02477	Human colon cancer
42	66.6	6.8	783	24	ABQ40788	Oligonucleotide fo
43	66.6	6.8	783	24	ABQ40789	Oligonucleotide fo
44	66	6.7	1172	24	ABQ31996	Oligonucleotide fo
45	66	6.7	1172	24	ABQ31997	Oligonucleotide fo

ALIGNMENTS

RESULT 1

AAV68668

ID AAV68668 standard; DNA; 979 BP.

XX AAV68668;

XX AC

XX DT

XX DT

XX 03-JUN-1999 (first entry)

XX DE Nucleotide sequence of the human Tango-81 gene.

XX KW Human; Tango-81; host cell; recombinant protein; antibody;

XX KW receptor; specific binding agent; probe; primer; hybridisation;

XX KW amplification; mutation; genetic mapping; ss.

XX OS Homo sapiens.

XX FH Key

XX CDS Location/Qualifiers

FT 58..840

FT /tag= a

FT /product= "Tango-81"

XX WO9906427-A1.

XX PD 11-FEB-1999.

XX PF 04-AUG-1998; 98WO-US16241.

XX PR 04-AUG-1997; 97US-0054645.

XX PA (MILL-) MILLENNIUM BIOTHEAPEUTICS INC.

XX PI McCarthy SA;

CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX

SQ Sequence 1041 BP; 145 A; 356 C; 402 G; 138 T; 0 other;

Query Match 98.2%; Score 961.4; DB 22; Length 1041;

Best Local Similarity 99.4%; Pred. No. 2.3e-140; Mismatches 1; Indels 0; Gaps 0;
Matches 959; Conservative 5;

QY 15 GCCAGCAGTCCGCCSGYMGRRRCGGGCTCGCTGGGGCAGCATGGCGGGTCCCGCT 74
DB 1004 GCCAGCAGTCCGCCSGYMGRRRCGGGCTCGCTGGGGCAGCATGGCGGGTCCCGCT 945
QY 75 GCTCTGGGGCGGGCGGGCGGCTGGGCTTTTGGTGTCTGTCTGTCTGTCTGTCTGT 134
DB 944 GCTCTGGGGCGGGCGGGCGGCTTTTGGTGTCTGTCTGTCTGTCTGTCTGTCTGT 885
QY 135 TCGGCGCGCGCGCGCTCTCTCGCGCGCGCGGTAAGAGAGCGCGCGCTTAAGCGAGC 194
DB 884 TCGGCGCGCGCGCGCTCTCTCGCGCGCGCGGTAAGAGAGCGCGCGCTTAAGCGAGC 825
QY 195 GTCTCGCGCGCTGTGAGTGTGCGCTCTCTCGCGCGCTCGCGCGCTCGAGTGTGCGCG 254
DB 824 GTCTCGCGCGCTGTGAGTGTGCGCTCTCTCGCGCGCTCGCGCGCTCGAGTGTGCGCG 765
QY 255 TGAGCGCGCGCGCGCTGTGAGTGTGCGCTCTCTCGCGCGCTCGCGCGCTCGAGTGTGCGCG 314
DB 764 TGAGCGCGCGCGCGCTGTGAGTGTGCGCTCTCTCGCGCGCTCGCGCGCTCGAGTGTGCGCG 705
QY 315 ACCTCAGGAGCGCGCGCGCGCGCGAGGCGCTGAGGATCAGAGCGCGCGCGCTCT 374
DB 704 ACCTCAGGAGCGCGCGCGCGCGCGAGGCGCTGAGGATCAGAGCGCGCGCGCTCT 645
QY 375 GCGCAGCTGTGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTGCGCTTGA 434
DB 644 GCGCAGCTGTGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTGCGCTTGA 585
QY 435 CGAGCGCGCGCGCGCTGTGAGTGTGCGCTCTCTCGCGCGCTCGCGCGCTCGAGTGTGCGCG 494
DB 584 CGAGCGCGCGCGCGCTGTGAGTGTGCGCTCTCTCGCGCGCTCGCGCGCTCGAGTGTGCGCG 525
QY 495 CCGTCCCGCGCTAGCAGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 554
DB 524 CCGTCCCGCGCTAGCAGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 465
QY 555 CCGCGCGCGCGCTGTGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 614
DB 464 CCGCGCGCGCGCTGTGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405
QY 615 GACACCG 674
DB 404 GACACCG 345
QY 675 CCGGCGCTCCGAGGGGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 734
DB 344 CCGGCGCTCCGAGGGGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 285
QY 735 GGGCTGTGAGTGTGCG 794
DB 284 GGGCTGTGAGTGTGCG 225

QY 795 GACCCCG 854
DB 224 GACCCCG 165
QY 855 CCGTGCACCTTGGGACCCAGAGTGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 914
DB 164 CCGTGCACCTTGGGACCCAGAGTGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105
QY 915 AGCAGCTCCAGAGCACTTACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 974
DB 104 AGCAGCTCCAGAGCACTTACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 45
QY 975 CTGGC 979
DB 44 CTGGC 40
RESULT 4
AAI5948/c
ID AAI5948 standard; cDNA; 1041 BP.
XX AC AAI5948;
XX AC AC
DT 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 3937.
DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX Homo sapiens.
OS
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR P-PSDB; AAM40792.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 3937; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38442-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

vector, an antibody against the secreted protein, a recombinant host cell that expresses the secreted protein and a method of identifying a binding partner of the secreted protein. The nucleic acid and protein are used to prevent, diagnose, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. Many other diseases and disorders are listed in the specification. The polypeptides can also be used to aid wound healing in an epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The present sequence encodes a novel human secreted protein of the invention.

Sequence 1050 BP: 172 A: 398 C: 344 G: 136 T: 0 other:

Sequence 1050 BP; 172 A; 398 C; 344 G; 136 T; 0 other:

Query Match 93.2%; Score 912.8; DB 24; Length 1050;
Best Local Similarity 98.9%; Pred. No. 7.2e-133;
Matches 955; Conservative 5; Mismatches 2; Indels 4;

Qy	14	GGCAGCAGTCCCGCGSSYMCGRRRCCCGGCTCGCTGGGGCAGCATGGCGGGTCCGCCG	73
Db	11	GGCAGCAGTCCCGCGCCCTCCGAGCCCGGCTCGCTGGGCGACATGGGGGTCGCCG	70
Qy	74	TGCTCTGGGGCGCGGGCGGGGCGCTCGCGCTTTTGGTGTCTGTCTGTCTGGCCTGT	133
Db	71	TGCTCTGGGGCGCGGGCGGGGCGTGGCCCTTTTGGTGTCTGTCTGTCTGGCCTGT	130
Qy	134	TTGGCGCGCCCCCGGCTCTGCGCGCGCGCGGTAAAGAGCCCCCGCCTTAAGCGCAG	193
Db	131	TTGGCGCGCCCCCGGCTCTGCGCGCGCGCGGTAAAGAGAGCCCCCGCCTTAAGCGCAG	190
Qy	194	CGTCTCGGCCCTTGGCTGACACTGGCGCTCTCTCGCGCTTTCGGCGGTCAGTCCCCGAG	253
Db	191	CGTCTCGGCCCTTGGCT-AGACTGGGGCTCTCTCGCGCTTTCGGCGGTCAGTCCCCGAG	249
Qy	254	GTGAGCGCGGGGGGGTGCAGGAGCTGGCGCGGCGCTGGCGCATCTGCTGGAGCGG	313
Db	250	GTGAGCGCGGGGGGGTGC-A-GACTGGCGGGCGCTGGCGCATCTGCTGGA-GCGG	307
Qy	314	AACGTCAGGAGCGGGCGCGGGCGCAGAGAGCTGAGGATCAGCAGGCGCGGTCC	373
Db	308	AAGCTCAGGAGCGGGCGGGCGCGGGCGAGCGCA-GAGGCTGAGGATCAGCAGGCGCGGTCC	366
Qy	374	TGGCGCAGCTGCTCGCGCTGTGGGGCGCCCCCGCAACTCTATCGGCTCTGGGCTTG	433
Db	367	TGGCGCAGCTGCTCGCGCTGTGGGGCGCCCCCGCAACTCTGATCGGCTCTGGGCTGG	426
Qy	434	ACAGACCCCGACGGCGCTTGCAGCGCAGTCTGCTCGCGCTGTCTGTCGCGCGCGCTTG	493
Db	427	ACGACACCCCGACGGCGCTTGCAGCGCAGTCTGCTCGCGCTGTCTGTCGCGCGCGCTTG	486
Qy	494	ACCTTCGCGCCCTAGCAGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGCGGCTCCGAC	553
Db	487	ACCTTCGCGCCCTAGCAGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGCGGCTCCGAC	546
Qy	554	CCGGCCCCCGGTCTACGACGACGGCCCCCGGGCCCCGATGCTGAGGAGGACGGCAGC	613
Db	547	CCGGCCCCCGGTCTACGACGACGGCCCCCGGGCCCCGATGCTGAGGAGGACGGCAGC	606
Qy	614	AGACCCGACGTGGACCCCGACCTTGTGAGGTACTTGTGGACCGGATTTCTGGGGAA	673
Db	607	AGACCCGACGTGGACCCCGACCTTGTGAGGTACTTGTGGACCGGATTTCTGGGGAA	666
Qy	674	GGCGGACTCCGAGGGGTGGACGCCCGCGCGCGCTCCGCGCTGCCCGCACCGATG	733
Db	667	GGCGGACTCCGAGGGGTGGACGCCCGCGCGCGCTCCGCGCTGCCCGCACCGATG	726

CC producing (I) in large quantities. Cells containing (II) are used for
CC the detection of ligands or receptors corresponding to membrane or
CC secretory proteins and to screen small molecule novel pharmaceuticals.
CC Antibodies directed to (I) can be used for the detection, quantification
CC and purification of (I). Activities of (I) may include cytokine and cell
CC proliferation/differentiation function, immune stimulating or suppressing
CC activity, haematopoiesis regulating activity, tissue growth activity, haemostatic
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory
CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.
CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
CC inflammatory bowel disease and tumours. (I) and (II) can also be used for
CC wound healing, as nutritional sources or supplements e.g. as amino acid,
CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
CC processing and utilisation of dietary fat, protein, carbohydrate,
CC vitamins and minerals, to effect behavioural characteristics, to affect
CC appetite, and can act as antigens in vaccines to raise an immune response
CC to the protein or another material cross-reactive with the protein.
XX

SQ Sequence 780 BP; 79 A; 297 C; 294 G; 110 T; 0 other;

Query Match 79.3%; Score 776.8; DB 22; Length 780;

Best Local Similarity 99.7%; Pred. No. 7.5e-112;

Matches 778; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 58 ATGGCGGGTCTGCGGCTCTCTGGGGCGCGGGCGGGGGCGGCTTTTGGTGTCTG 117
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 ATGGCGGGTCTGCGGCTCTCTGGGGCGCGGGCGGGGGCGGCTTTTGGTGTCTG 60
QY 118 CTGCTGCTCGGCGCTGTTTGGGGCGCGCGCGCTCTGCGCGCGCGGTAAAGGAGCCC 177
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 CTGCTGCTCGGCGCTGTTTGGGGCGCGCGCGCTCTGCGCGCGCGGTAAAGGAGCCC 120
QY 178 CGCGGGCTAAGCGCAGCGTCTCGCGCTTGGCTGAGACTGGCGCTCTCGCGCTTCCGG 237
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 CGCGGGCTAAGCGCAGCGTCTCGCGCTTGGCTGAGACTGGCGCTCTCGCGCTTCCGG 180
QY 238 CGGTGAGTCCCGAGGTGAGCGGGCGGGCGGGTGCAGGAGTGGCGGGCGCTGGCG 297
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 CGGTGAGTCCCGAGGTGAGCGGGCGGGCGGGTGCAGGAGTGGCGGGCGCTGGCG 240
QY 298 CATCTGCTGAGGCGCAACCTCAGGAGCGGGCGGGCGCGAGCGCAGGAGGCTGAGGAT 357
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 CATCTGCTGAGGCGCAACCTCAGGAGCGGGCGGGCGCGAGCGCAGGAGGCTGAGGAT 300
QY 358 CAGCAGCGCGCGTCTGCGCGAGCTGCTGCGCGTCTGGGGCGCGCGCGCAACTCTGAT 417
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 CAGCAGCGCGCGTCTGCGCGAGCTGCTGCGCGTCTGGGGCGCGCGCGCAACTCTGAT 360
QY 418 CCGGCTTGGGCTTGGACGACGCGCGCGCGCGCGCGCTGCGAGCGAGTCTGCGCTCTG 477
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 CCGGCTTGGGCTTGGACGACGCGCGCGCGCGCGCGCTGCGAGCGAGTCTGCGCTCTG 420
QY 478 CTCGCGCGCGCTTGACCTGCGCGCTAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 537
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
421 CTCGCGCGCGCTTGACCTGCGCGCTAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 538 GCGCGCGCGCTCGACCGCGCGCGCGCGCGCTACGAGCGCGCGCGCGCGCGCGCGGATGCT 597
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
481 GCGCGCGCGCTCGACCGCGCGCGCGCGCGCTACGAGCGCGCGCGCGCGCGCGCGGATGCT 540
QY 598 GAGGAGCGCGCGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGATGCT 657
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
541 GAGGAGCGCGCGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGATGCT 600
QY 658 CGGATTTCTCGGGAGCGCGGACTCCGAGGGGGTGCAGCGCGCGCGCGCGCGCGCGCGCT 717
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
601 CGGATTTCTCGGGAGCGCGGACTCCGAGGGGGTGCAGCGCGCGCGCGCGCGCGCGCGCT 660
QY 718 GCCCGCGCACCATGTGGGCTCTGAGCTGCCCGCTGAGGGCGGTGCTGGGGGGCGCTGCTG 777
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
661 GCCCGCGCACCATGTGGGCTCTGAGCTGCCCGCTGAGGGCGGTGCTGGGGGGCGCTGCTG 720

QY 778 CGTGTGAAGCGCTAGAGACCGCGCGCGCGCGCGCGCGCTGCTGACGCGCGCTTGTGCACCC 837
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
721 CGTGTGAAGCGCTAGAGACCGCGCGCGCGCGCGCGCGCTGCTGACGCGCGCTTGTGCACCC 780

RESULT 11

AAI58162
ID AAI58162 standard; cDNA; 919 BP.
AC AAI58162;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 365.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.

XX Homo sapiens.

XX OS WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-052317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM39006.

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 365; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 919 BP; 122 A; 365 C; 308 G; 124 T; 0 other;

Query Match 74.5%; Score 729.4; DB 22; Length 919;

XX	15-JUL-2002	(first entry)
XX	Human secreted protein gene 49 genomic DNA fragment #2.	
XX	Human; ds; secreted protein; gene therapy; immunosuppressive;	
XX	antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;	
XX	vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;	
XX	virucide; fungicide; ophthalmological; autoimmune disease; neoplasm;	
XX	rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;	
XX	cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;	
XX	angiogenesis; nervous system disorder; Alzheimer's disease; infection;	
XX	ocular disorder; corneal infection; wound healing; skin aging;	
XX	epithelial cell proliferation; food additive.	
XX	Homo sapiens.	
XX	WO200226931-A2.	
XX	04-APR-2002.	
XX	24-SEP-2001; 2001WO-US99871.	
XX	25-SEP-2000; 2000US-234925P.	
XX	12-JAN-2001; 2001WO-US00911.	
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;	
XX	Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;	
XX	Endress GA, Mucenski M, Ebner R;	
XX	WPI; 2002-362489/39.	
XX	Novel 71 isolated secreted polypeptides and polynucleotides encoding	
XX	the polypeptides, useful for treating Huntington's disease, sepsis,	
XX	meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,	
XX	asthma -	
XX	Example 2; Page 1404-1406; 1478pp; English.	
XX	The invention relates to an isolated nucleic acid molecule (or its	
XX	fragment, homologue complement or allelic variant) encoding a human	
XX	secreted protein (and its fragment, domain, epitope, variant, secreted	
XX	form and species variant). Also included are a recombinant vector	
XX	comprising the nucleic acid, a recombinant host cell comprising the	
XX	vector, an antibody against the secreted protein, a recombinant host cell	
XX	that expresses the secreted protein and a method of identifying a binding	
XX	partner of the secreted protein. The nucleic acid and protein are used to	
XX	prevent, diagnose, treat or ameliorate a medical condition in e.g.	
XX	humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep	
XX	for example autoimmune diseases e.g. rheumatoid arthritis,	
XX	hyperproliferative disorders e.g. neoplasms of the breast or liver,	
XX	cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders	
XX	e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.	
XX	Alzheimer's disease, infections caused by bacteria, viruses and fungi and	
XX	ocular disorders e.g. corneal infection. Many other diseases and	
XX	disorders are listed in the specification. The polypeptides can also be	
XX	used to aid wound healing an epithelial cell proliferation, to prevent	
XX	skin aging due to sunburn, to maintain organs before transplantation, for	
XX	supporting cell culture of primary tissues, to regenerate tissues and in	
XX	chemotaxis. The polypeptides can also be used as a food additive or	
XX	preservative to increase or decrease storage capabilities. The present	
XX	sequence represents a ds DNA fragment of the gene for a novel human	
XX	secreted protein of the invention.	

Query Match 48.4%; Score 474.2; DB 24; Length 4503;
Best Local Similarity 99.4%; Pred. No. 3.3e-65;
Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0

ov 170 AGAGACCGCGCGCCTAAGCGCAGCGTTCGCGCCCTGGCTGAGACTGGCGGCTCCGCC 229

```
Db 3254 AGAGCCCGCGGCTAAGCAGCGTCTCCGCCCTTGCTGAGACTGCGCTCTCTCGCC 3313
QY 230 GCTTCCGGCGGTCACTGCCCCAGGTGAGCGCGCGGGCGGTGAGGAGCTGGCGCGGG 289
Db 3314 GCTTCCGGCGGTCACTGCCCCAGGTGAGCGCGCGGGCGGTGAGGAGCTGGCGCGGG 3373
QY 290 CCGTGGCGCATCTGCTGGAGCGCCGAACGTCAGAGCGGCGCGGGCCGAGGCGCAGGAGG 349
Db 3374 CGCTGGCGCATCTGCTGGAGCGCCGAACGTCAGAGCGGCGCGGGCCGAGGCGCAGGAGG 3433
QY 350 CTGAGATCAGACGCGCGCTCTGCGCAGCTGCTGGCGCTCTGGCGCGCGCGCCCGCA 409
Db 3434 CTGAGATCAGACGCGCGCTCTGCGCAGCTGCTGGCGCTCTGGCGCGCGCGCCCGCA 3493
QY 410 ACTGATCCCGCTCTGGGCTTGGAGCGACGACCCGCGCGCTGCGAGCGAGCTCGTCTC 469
Db 3494 ACTGATCCCGCTCTGGGCTTGGAGCGACGACCCGCGCGCTGCGAGCGAGCTCGTCTC 3553
QY 470 GCGCTCTGCTCGCGCGCGCTTGACCTGCGCGCTAGCAGCCAGCTGTCTCCCGCGC 529
Db 3554 GCGCTCTGCTCGCGCGCGCTTGACCTGCGCGCTAGCAGCCAGCTGTCTCCCGCGC 3613
QY 530 CGGTCCCGCGCGCGCTTCCGACCCCGCGCGCTTACGACGACGCGCGCGCGCGCC 589
Db 3614 CGGTCCCGCGCGCGCTTCCGACCCCGCGCGCTTACGACGACGCGCGCGCGCGCC 3673
QY 590 CGGATGCTGAGGAGCGAGCGAGACACCCGCGAGCTGAGCCCGCGAGCTGTGAGGTAC 648
Db 3674 CGGATGCTGAGGAGCGAGCGAGACACCCGCGAGCTGAGCCCGCGAGCTGTGAGGTAC 3732

RESULT 14
ABQ31154/c
ID ABQ31154 standard; DNA; 652 BP.
XX
AC ABQ31154;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17745.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN W0200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
```

```
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP-s); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ154121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
```

SEQ Sequence 652 BP; 103 A; 81 C; 233 G; 235 T; 0 other;

```
Query Match 30.8%; Score 301.6; DB 24; Length 652;
Best Local Similarity 77.1%; Pred. No. 1.9e-38;
Matches 367; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 173 AGCCCGCGCGCTAAGCGAGCGTCTCGGCCCTTGCTGAGACTGCGCTCTCTCGCGCT 232
Db 511 AACC CGCGCGCTAAGCGAGCGTCTCGGCCCTTAACTAAACTAACGCTCTCTCGCGCT 452
QY 233 TCCGCGCGCTGAGTCCCGCGAGGTGAGCGCGCGCGCGCTGAGGAGCTGCGCGCGCG 292
Db 451 TCCGAGGATCAATACCCCGGAAATAAAGACGAAACGATACAAAACTAACGCGAAGCG 392
QY 293 TGGCGCATCTGCTGGAGCGCGAAGCTCAGAGCGCGCGCGCGCGCGAGGAGGCTG 352
Db 391 TAACGCATCTACTTAAACCGGAACGCTCAAAACGAAACGCGAACCAGCAAAACTA 332
QY 353 AGGATCAGACGCGCGCTCTCTGCGCGAGCTGCTGCGCGCTGCGCGCGCGCGCGCG 412
Db 331 AAAATCAACAAACGCGCTCTCTGAGCACTACTACTGCGCTCTAAACCGCGCGCGCACT 272
QY 413 CTGATCCCGCTCTGCGCTTGGAGCAGACGACCGCGCGCGCGCTGAGGAGCTGCTCGCG 472
Db 271 CTATTCGAGCTCTAAACCTAAACGAGACCGCGCGCGCGCTTAAACGCAACTCTCGCG 212
QY 473 CTCTGCTCCGCGCGCGCTTGAACCTGCGCGCTAGCAGCCAGCTGTGCTCCCGCGCGCG 532
Db 211 CTCTACTCCGCGCGCGCTTAAACCTACCGCGCTCGCAACCACTTATCCCGCGCGCG 152
QY 533 TCCCGCGCGCGCGCTCCGAGCCCGCGCGCGCGCTTACGAGACGCGCGCGCGCGCG 592
Db 151 TCCCGCGCGCGAGCTCCGAGCCCGCGCGCGCGCTTACGAGACGCGCGCGCGCGCG 92
QY 593 ATGCTGAGGAGCGCGCGAGCAGACACCGCGAGTGTGAGCGCTGTGAGGTAC 648
Db 91 ATACTAAAAAACAACGCAACCGAATACCCGAGTAAACCCCGCACTATTAAATAAC 36
```

RESULT 15
ABQ31155

ID ABQ31155 standard; DNA; 652 BP.

XX ABQ31155;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17746.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX WO200218632-A2.
XX PD 07-MAR-2002.
XX PF 01-SEP-2001; 2001WO-EP10074.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PR 05-SEP-2000; 2000DE-104543.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX DR WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA -
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridisation to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridised to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (i) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc., particularly by detecting mutations or single nucleotide
XX polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX method for determining the degree of cytosine methylation described in
XX the disclosure of the invention.
XX
XX Sequence 652 BP; 235 A; 233 C; 81 G; 103 T; 0 other;

QY 533 TCCCGCGCGGCGCTCGGACCCCGCGCGCGGCTCTAGACGAGCGCGCGCGCGCGG 592
Db 502 TCCCGCGCGGAGCGCTCCGACCCCGCGCGCGGATCTACGACGACCGCGCGCGG 561
QY 593 ATGCTGAGGAGGCGACGAGACACCCGACGCTGGACCCCGGAGCTGTGAGGTAC 648
Db 562 ATACTAAACAAACAGACGAAACCCGACGTAACCCCGAACTATTAAATAC 617

Search completed: April 12, 2003, 17:22:03
Job time : 292 secs

Query Match 30.8%; Score 301.6; DB 24; Length 652;
Best Local Similarity 77.1%; Pred. No. 1.9e-38;
Matches 367; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 173 AGCCCGCGGCGCTAAGCGCAGCGTCTCCGCCCTTGGCTGAGACTGGCGCTCCCTCGCGCT 232
Db 142 AACCCCGCGAGCTTAACGCAACGCTCTCCGCCCTTAACTAAACCTAACGCTCTCTCGCGCT 201
QY 233 TCCCGCGGCTCAGTCCCGGAGTGAGCGGGGGGGTGCAGGAGTGGCGGGCGC 292
Db 202 TCCGACGATCATACCCCGAATAAAGAGGAAACGATACAAAACCTAACGCGAAGC 261
QY 293 TGGCGCATCTGCTGGAGCGCGCAAGCTCAGGAGCGGGCGCGGCGGAGCGAGGAGGCTG 352
Db 262 TAACGCATCTACTAAAAACCGACGCTCAAAAACGACGCAACCGAACAACCACTA 321
QY 353 AGGATCAGCAGGCGCGGCTCTGGCGGAGCTGTGGCGGCTGGGGGCGCGCGCAACT 412
Db 322 AAAATCAACAACCGCGGCTCTTACGCAACTACTACGCGCTTAAACGACCGCGCGCAACT 381
QY 413 CTGATCCGGCTCTGGGCTTGGACGACGACCGCGCTGCGCGAGCTGCTCGCG 472
Db 382 CTATCCGACTCTAAACCTAAACGACGACCGCGCGGCGCTTACACGCAACTCGCTCGG 441
QY 473 CTCGTCTCCGCGCGCGCTTACCGCTTACCGCTTACGAGCGCGCTTGTCCCGCGCGCG 532
Db 442 CTCCTACTCCGCGCGCGCTTAAACCGCTTACCGCTTACCGCTTACCGCACTTATCCCGCGCGG 501

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:08:39 ; Search time 37 Seconds
(without alignments)
936.356 Million cell updates/sec

Title: us-09-803-589-6

Perfect score: 1319

Sequence: 1 MAGSPLWPGAGGVGLLV.....RVKRLTPAPQVPARRLLPP 260

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_101002.*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
 - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	260	20	AAW84597
2	1319	100.0	260	21	AAY71959
3	1319	100.0	260	22	AAW39005
4	1319	100.0	260	22	AAW88596
5	1105.5	83.8	223	22	AAW39006
6	1019	77.3	258	21	AAY71960
7	1015	77.0	262	21	AAY71962
8	308	23.4	148	22	AAE06088
9	308	23.4	148	23	ABG33910
10	308	23.4	149	21	AAY87111

11	308	23.4	166	21	AAY87199	Human secreted pro
12	308	23.4	166	22	ABG06176	Human gene 48 enco
13	308	23.4	166	23	ABG34000	Human secreted pro
14	259	19.6	69	21	AAY71961	Rat pituitary horm
15	136.5	10.3	388	23	ABG60300	Lymphoma associate
16	136.5	10.3	388	23	ABG09271	G protein-coupled
17	136.5	10.3	451	22	AAU68528	Human novel cytoke
18	134.5	10.2	317	22	AAW40791	Human polypeptide
19	134.5	10.2	317	22	AAW40792	Human polypeptide
20	124	9.4	720	23	AAE20839	Human gene 7 encod
21	124	9.4	766	23	AAE20838	Human gene 7 encod
22	124	9.4	973	23	ABG64875	Human albumin fusi
23	124	9.4	973	23	ABG64875	Human gene 7 encod
24	123.5	9.4	597	23	ABG20799	Human secretory po
25	123	9.3	251	22	AAU04906	Micromonospora eve
26	121.5	9.2	481	22	ABG05355	Novel human diago
27	119.5	9.1	248	15	ABG46912	GGF segment E. Ho
28	119.5	9.1	248	15	AAW55853	GGF segment E. Ho
29	119.5	9.1	248	17	AAW09358	Human glial growth
30	119.5	9.1	248	17	AAW09372	Human glial growth
31	119.5	9.1	248	17	AAW09372	Human glial growth
32	119.5	9.1	248	20	AAW87445	Human glial growth
33	119.5	9.1	422	15	AAW55854	Human neurogulin g
34	119.5	9.1	422	15	AAW46923	GGF-II encoded by
35	119.5	9.1	422	16	AAW67258	GGF-II encoded by
36	119.5	9.1	422	17	AAW09371	Human glial cell g
37	119.5	9.1	422	17	AAW09372	Human neurogulin g
38	119.5	9.1	422	17	AAW09372	Human GGF2. Homo
39	119.5	9.1	422	17	AAW87466	Glial growth facto
40	119.5	9.1	422	17	AAW87467	Glial growth facto
41	119.5	9.1	422	17	AAW86628	Glial growth facto
42	119.5	9.1	422	22	AAW67901	Mature hGGF2. Hom
43	119.5	9.1	422	22	AAW67901	Human neurogulin g
44	119	9.0	404	19	AAW68408	Aujesky's disease
45	119	9.0	421	22	ABW71817	Drosophila melanog

ALIGNMENTS

RESULT 1

AAW84597
ID AAW84597 standard; Protein; 260 AA.

XX AAW84597:

XX AC

XX 03-JUN-1999 (first entry)

XX Amino acid sequence of the human Tango-81 protein.

XX Human; Tango-81; host cell; recombinant protein; antibody;

XX receptor; specific binding agent; probe; primer; hybridisation;

XX amplification; mutation; genetic mapping.

XX Homo sapiens.

XX WO9906427-A1.

XX 11-FEB-1999.

XX 04-AUG-1998; 98WO-US16241.

XX 04-AUG-1997; 97US-0054645.

XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

XX McCarthy SA;

XX WPI; 1999-153693/13.

XX N-PSDB; AAW68668.

XX New nucleic acid encoding human Tango-78, -79 and -81 proteins -

XX useful for diagnosis and treatment of Tango-associated diseases

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX Claim 1; Fig 3; 67pp; English.

PS This is the amino acid sequence of the human Tango-81 protein

XX used in the method of the invention. Host cells containing

CC the Tango protein are used to produce recombinant proteins for

CC raising antibodies. It is also used in identifying specific

CC binding agents (including cognate receptors), which can be used to

CC determine amounts of recombinant protein in cells or

CC therapeutically. Antibodies or other specific binding agents, are

CC used to detect recombinant proteins and fragments of the Tango

CC nucleotide sequence can be used as probes or primers for detecting

CC the Tango gene, specifically mRNA, in usual hybridisation or

CC amplification assays. These assays are used for diagnosis of

CC diseases associated with abnormal expression of Tango proteins,

CC e.g. detecting mutations in the Tango gene. Fragments of the Tango

CC nucleic acid sequence are also used for genetic mapping and

CC chromosome identification, and as antisense, ribozyme or

CC triplex-forming therapeutics. Antibodies may also be used to generate

XX anti-idiotypic antibodies.

XX Sequence 260 AA;

SQ Query Match 100.0%; Score 1319; DB 20; Length 260;

Best Local Similarity 100.0%; Pred. No. 2.9e-104;

Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLMGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60

DB 1 MAGSPLLMGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60

QY 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120

DB 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120

QY 121 PALGLDDDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPVPPVYDDGAPGPD 180

DB 121 PALGLDDDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPVPPVYDDGAPGPD 180

QY 181 EAGDETTPDPELLRYLGRILASADSEGVAAAPRRRLRAADHDVGSLEPPEGVLGALL 240

DB 181 EAGDETTPDPELLRYLGRILASADSEGVAAAPRRRLRAADHDVGSLEPPEGVLGALL 240

QY 241 RVKRLTAPQVPAARLLPP 260

DB 241 RVKRLTAPQVPAARLLPP 260

RESULT 2

AAV71959

ID AAV71959 standard; Protein; 260 AA.

XX AAV71959;

XX 26-MAR-2001 (first entry)

XX Human pituitary hormone, pituitrone.

XX Human; pituitary; pituitrone; therapy; immune disorder; anaemia;

KW Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;

KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;

KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;

KW SLE; hyperproliferative disorder; gene therapy; neoplasm;

KW infectious disease; immunomodulatory; cytostatic; antimicrobial.

OS Homo sapiens.

XX Key Location/Qualifiers

EH 1..33 /label= signal_peptide

FT 34..260 /label= Mature_human_pituitrone

FT 36..44

FT Region /note= "Highly immunogenic"

FT 55..68

FT Region /note= "Highly immunogenic"

FT 116..121

FT Region /note= "Highly immunogenic"

FT 123..131

FT Region /note= "Highly immunogenic"

FT 164..182

FT Region /note= "Highly immunogenic"

FT 185..193

FT Region /note= "Highly immunogenic"

FT 206..211

FT Region /note= "Highly immunogenic"

FT 225..233

FT Region /note= "Highly immunogenic"

FT 253..258

FT Region /note= "Highly immunogenic"

XX WO200066778-A1.

XX 09-NOV-2000.

XX 27-APR-2000; 2000WO-US11211.

XX 30-APR-1999; 99US-0131966.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Ni J;

PI WPI; 2000-687547/67.

XX N-PSDB; AAD02075.

DR Nucleic acids encoding a novel pituitary hormone (pituitrone), useful

DR for diagnosing, preventing and treating e.g. immune disorders,

PT hyperproliferative disorders and blood coagulation disorders -

XX Claim 11; Fig 1; 277pp; English.

XX The present sequence is a human pituitary hormone, pituitrone (clone

CC HKGL36). Pituitrone is highly expressed in pituitary

CC gland and is also expressed in brain tissues, spinal cord and kidney.

CC It may be used as antigens in the production of antibodies against

CC pituitrone and in assays to identify modulators. Pituitrone cDNAs are

CC also useful in gene therapy. Pituitrone may be used in the prevention,

CC treatment and diagnosis of diseases associated with inappropriate

CC pituitrone expression. It may be useful in treating disorders related

CC to reproductive and renal system, immune disorders, disorders of

CC haematopoietic cells (e.g., anaemia, Digeorge syndrome, ataxia

CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation

CC disorders, autoimmune disorders (e.g. Addison's disease, multiple

CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative

CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious

CC diseases.

XX Sequence 260 AA;

SQ Query Match 100.0%; Score 1319; DB 21; Length 260;

Best Local Similarity 100.0%; Pred. No. 2.9e-104;

Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGSPLLMGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60

DB 1 MAGSPLLMGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETGAPRRFR 60

QY 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120

DB 61 RSVPRGEAAGAVQELARALAHLEAEQERARAEAEQEDQARVLAQLLRVWGAPRNSD 120

QY 121 PALGLDDDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPVPPVYDDGAPGPD 180

DB 121 PALGLDDDDPDAPAAQALARALLRARLDPAALAAQLVPAPVPAALRPVPPVYDDGAPGPD 180

Qy	181	EEAGDETPDVP	PELLYLLGRIL	LAGSADSEGV	AAARRLRADHD	VGSSELP	PEGVLGALL	240
Db	181	EEAGDETPDVP	PELLYLLGRIL	LAGSADSEGV	AAARRLRADHD	VGSSELP	PEGVLGALL	240
Qy	241	RVKRLETPAPQ	VPARLLPP					260
Db	241	RVKRLETPAPQ	VPARLLPP					260

xx	SQ	Sequence	260 AA;
		Query Match	100.0%; Score 1319; DB 22; Length 260;
		Best Local Similarity	100.0%; Pred. No. 2.9e-104;
		Matches 260; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	MAGSPLLMGPRAGVGGLVLLILGLFRPPDPAALCAARPYKEPRGLSAAAPPLAETGAPRRR	60
Dd	1	MAGSPLLMGPRAGVGGLVLLILGLFRPPDPAALCAARPYKEPRGLSAAAPPLAETGAPRRR	60
Qy	61	RSDVRGEAGA VQELARALAHLLAEERQERARAFAQAEDQQARVLAQLLRVWGAPRNSD	120
Dd	61	RSDVRGEAGA VQELARALAHLLAEERQERARAFAQAEDQQARVLAQLLRVWGAPRNSD	120
Qy	121	PALGLDDDDPAPAQAQLARALLRARLDPAALAQAQLVPAPVAALRPDPVYDDGPAGPDA	180
Dd	121	PALGLDDDDPAPAQAQLARALLRARLDPAALAQAQLVPAPVAALRPDPVYDDGPAGPDA	180
Qy	181	EEAGDETPDVDPPELLRYLLGRILAGSADSEGVAAAPRLRLRAADHDVGSPELVGLALL	240
Dd	181	EEAGDETPDVDPPELLRYLLGRILAGSADSEGVAAAPRLRLRAADHDVGSPELVGLALL	240
Qy	241	RVKRLTAPAQVPARRLLPP	260
Dd	241	RVKRLTAPAQVPARRLLPP	260
RESULT	4		
AAB88596	ID	AAB88596 standard; Protein; 260 AA.	
XX	XX	AAB88596;	
XX	AC		
XX	XX		
DT	04-JUN-2001	(first entry)	
XX	XX	Human hydrophobic domain containing protein clone HP10756 #100.	
XX	XX	Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;	
KW	KW	antianaemic; vulnery; antituler; osteopathic; anti-inflammatory;	
KW	KW	cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;	
KW	KW	HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;	
KW	KW	inflammatory bowel disease; nutritional supplement; appetite; vaccine;	
KW	KW	behavioural characteristic; immune response.	
-XX	-XX		
OS	OS	Homo sapiens.	
XX	XX	WO200112660-A2.	
PN	PN	22-FEB-2001.	
XX	XX		
PD	PD	10-AUG-2000; 2000WO-JP05356.	
XX	XX		
PF	PF	17-AUG-1999; 99JP-0230344.	
XX	XX	07-SEP-1999; 99JP-0252551.	
PR	PR	01-OCT-1999; 99JP-0281132.	
PR	PR	22-OCT-1999; 99JP-0301624.	
PR	PR	04-NOV-1999; 99JP-0313877.	
XX	XX	(SAGA) SAGAMI CHEM RES CENT.	
PA	PA	(PROT-) PROTEGENE INC.	
XX	XX		
PI	PI	Kato S, Kimura T;	
XX	XX		
DR	DR	WFI; 2001-160059/16.	
DR	DR	N-PSDB; AAF94486.	
XX	XX		
PT	PT	Human proteins with hydrophobic domains and the DNAs which encode them	
PT	PT	are useful for treating autoimmune disorders, burns and tumors and for	
PT	PT	screening novel pharmaceuticals -	
XX	XX	Claim 1; Page 422-423; 518pp; English.	
PS	PS		
XX	XX		

CC AAF94417 to AAF94516 encode the human proteins given in AAB88557 to
CC AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
CC anti-HIV, neuroprotective, antianemic, vulnerary, antitumor,
CC osteopathic, anti-inflammatory and cytostatic activities, and can be
CC used in gene therapy. (I) can be used as pharmaceuticals and as antigens
CC to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
CC probes for genetic diagnosis and gene sources for gene therapy or for
CC producing (I) in large quantities. Cells containing (II) are used for
CC the detection of ligands or receptors corresponding to membrane or
CC secretory proteins and to screen small molecule novel pharmaceuticals.
CC Antibodies directed to (I) can be used for the detection, quantification
CC and purification of (I). Activities of (I) may include cytokine and cell
CC proliferation/differentiation function, immune stimulating or suppressing
CC activity, haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory
CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.
CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
CC inflammatory bowel disease and tumours. (I) and (II) can also be used for
CC wound healing, as nutritional sources or supplements e.g. as amino acid,
CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
CC processing and utilisation of dietary fat, protein, carbohydrate,
CC vitamins and minerals, to effect behavioural characteristics, to affect
CC appetite, and can act as antigens in vaccines to raise an immune response
CC to the protein or another material cross-reactive with the protein.
XX
SQ

Query Match 100.0%; Score 1319; DB 22; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.9e-104;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGSPLLWGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLAETGAPRRFR 60
DB 1 MAGSPLLWGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLAETGAPRRFR 60
QY 61 RSVPRGEAAGVQELARALAHLEAEERQERARAEAEQDQARVLAQLLRVWGAPRNSD 120
DB 61 RSVPRGEAAGVQELARALAHLEAEERQERARAEAEQDQARVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPAPAAQALARALLRDLRDLPAALAAQLVPAPVPAALRPPVYDDGAPGPA 180
DB 121 PALGLDDDDPAPAAQALARALLRDLRDLPAALAAQLVPAPVPAALRPPVYDDGAPGPA 180
QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRRAADHDVGSLEPPEGVIGALL 240
DB 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRRAADHDVGSLEPPEGVIGALL 240
QY 241 RVKRLTAPQVPARRLLPP 260
DB 241 RVKRLTAPQVPARRLLPP 260

RESULT 5
AAM39006
ID AAM39006 standard; Protein: 223 AA.
XX
AC AAM39006;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2151.
XX
KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.

XX 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
DR N-PSDB; AAI58162.
XX WPI; 2001-442253/47.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2151; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AA42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 223 AA;
Query Match 83.8%; Score 1105.5; DB 22; Length 223;
Best Local Similarity 85.8%; Pred. No. 3.4e-86;
Matches 223; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 MAGSPLLWGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLAETGAPRRFR 60
DB 1 MAGSPLLWGPRAGGVLVLLGLFRPPPPALCARPVKEPRGLSAASPPLAETGAPRRFR 60
QY 61 RSVPRGEAAGVQELARALAHLEAEERQERARAEAEQDQARVLAQLLRVWGAPRNSD 120
DB 61 RSVPRGEAAGVQELARALAHLEAEERQERARAEAEQDQARVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPAPAAQALARALLRDLRDLPAALAAQLVPAPVPAALRPPVYDDGAPGPA 180
DB 121 PALGLDDDDPAPAAQALARALLRDLRDLPAALAAQLVPAPVPAALRPPVYDDGAPGPA 180
QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRRAADHDVGSLEPPEGVIGALL 240
DB 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRRAADHDVGSLEPPEGVIGALL 240
QY 241 RVKRLTAPQVPARRLLPP 260
DB 204 RVKRLTAPQVPARRLLPP 223
RESULT 6

AAV71960
ID AAV71960 standard; Protein; 258 AA.
XX
AC AAV71960;
XX
DT 26-MAR-2001 (first entry)
XX
DE Mouse pituitary hormone, pituitrone.
XX
DE
XX
KW Mouse; pituitary; pituitrone; therapy; immune disorder; anaemia;
KW Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;
KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW SLE; hyperproliferative disorder; gene therapy; neoplasm;
KW Infectious disease; immunomodulatory; cytostatic; antimicrobial.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Peptide 1..34
FT /label= Signal_peptide
FT /note= "An alternative signal sequence referred in
FT page 11"
FT Peptide 1..33
FT /label= Signal_peptide
FT Protein 35..258
FT /note= "Mature mouse pituitrone"
FT /note= "An alternative mature protein referred in
FT page 11"
FT Protein 34..258
FT /note= "Mature mouse pituitrone"
XX
PN W0200066778-A1.
XX
XX 09-NOV-2000.
XX
XX 27-APR-2000; 2000WO-US11211.
XX
XX 30-APR-1999; 99US-0131966.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J;
XX
XX WPI; 2000-687547/67.
DR N-PSDB; AAD02076.
XX
XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
PT for diagnosing, preventing and treating e.g. immune disorders,
PT hyperproliferative disorders and blood coagulation disorders -
XX
XX Claim 11; Fig 2; 277pp; English.
PS
XX The present sequence is a mouse pituitary hormone, pituitrone.
CC Pituitrone is highly expressed in pituitary gland and is
CC also expressed in brain tissues, spinal cord and kidney. It may be
CC used as antigens in the production of antibodies against pituitrone
CC and in assays to identify modulators. Pituitrone cDNAs are
CC also useful in gene therapy. Pituitrone may be used in the prevention,
CC treatment and diagnosis of diseases associated with inappropriate
CC pituitrone expression. It may be useful in treating disorders related
CC to reproductive and renal system, immune disorders, disorders of
CC haematopoietic cells (e.g., anaemia, Digeorge syndrome, ataxia
CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation
CC disorders, autoimmune disorders (e.g. Addison's disease, multiple
CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
XX diseases.
SQ Sequence 258 AA;

Query Match 77.3%; Score 1019; DB 21; Length 258;
Best Local Similarity 79.2%; Pred. No. 9.2e-79;

Matches 206; Conservative 13; Mismatches 39; Indels 2; Gaps 1;
QY 1 MAGSPLWPGRAGGVLVLLGLFRPPFALCARPVKEPRGLSAAAPPLAETGAPRRFR 60
DB 1 MAGSPLLLRAAGGGVILVLLGLLRLPTLSARPVKEPRSLARSAPLVTSTPLRLR 60
QY 61 RSVPRGEAAGVQELARALAHLLLEAEQERARAEAEQEQARVLAQLLRVWCAPNRSD 120
DB 61 RAVPRGEAAGVQELARAVATVLEAEQERARAEAEQEQARVLAHLPPWGSPRASD 120
QY 121 PALGLDDDDPAPAAQALRALRLRDLPAALAAQLVPAPVPAALRPPRPVYDDGPGAPDA 180
DB 121 PPLAPDDDDPAPAAQALRALRLRDLPAALAAQLVPA--PAAAPRPVYDDGPTGPDV 178
QY 181 EAGDETPDVPDPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVSELPEGVGLALL 240
DB 179 EDAGDETPDVPDPELLRYLLGRILTSSEPEAAAPRRLRSVDQDLGPEVPPENVLGLI 238
QY 241 RVKLETPAPQVPARRLLPP 260
DB 239 RVKLENPSQAPARRLLPP 258
RESULT 7
AAV71962
ID AAV71962 standard; Protein; 262 AA.
XX
AC AAV71962;
XX
DT 26-MAR-2001 (first entry)
XX
DE Consensus sequence of human, mouse and rat pituitrone.
XX
KW Rat; human; mouse; pituitary; pituitrone; therapy; immune disorder;
KW Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;
KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW SLE; hyperproliferative disorder; gene therapy; neoplasm; anaemia;
KW Infectious disease; immunomodulatory; cytostatic; antimicrobial.
XX
OS Rattus sp.
OS Mus sp.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 8..12
FT /label= Unknown
FT Misc-difference 17
FT /label= Unknown
FT Misc-difference 26
FT /label= Unknown
FT Misc-difference 28
FT /label= Unknown
FT Misc-difference 31
FT /label= Unknown
FT Misc-difference 33
FT /label= Unknown
FT Misc-difference 42
FT /label= Unknown
FT Misc-difference 46
FT /label= Unknown
FT Misc-difference 48
FT /label= Unknown
FT Misc-difference 51
FT /label= Unknown
FT Misc-difference 54
FT /label= Unknown
FT Misc-difference 55
FT /label= Unknown
FT Misc-difference 57
FT /label= Unknown
FT Misc-difference 59
FT /label= Unknown

FT Misc-difference 62 /label= Unknown
FT Misc-difference 79 /label= Unknown
FT Misc-difference 81 /label= Unknown
FT Misc-difference 82 /label= Unknown
FT Misc-difference 108 /label= Unknown
FT Misc-difference 111 /label= Unknown
FT Misc-difference 112 /label= Unknown
FT Misc-difference 115 /label= Unknown
FT Misc-difference 118 /label= Unknown
FT Misc-difference 122 /label= Unknown
FT Misc-difference 124 /label= Unknown
FT Misc-difference 125 /label= Unknown
FT Misc-difference 159 /label= Unknown
FT /label= Unknown
FT /note= "Xaa equals any amino acid residue or no
FT amino acid residue"
FT /label= Unknown
FT /note= "Xaa equals any amino acid residue or no
FT amino acid residue"
FT /label= Unknown
FT /label= Unknown
FT Misc-difference 176 /label= Unknown
FT Misc-difference 180 /label= Unknown
FT Misc-difference 182 /label= Unknown
FT Misc-difference 207 /label= Unknown
FT /label= Unknown
FT /note= "Xaa equals any amino acid residue or no
FT amino acid residue"
FT /label= Unknown
FT /note= "Xaa equals any amino acid residue or no
FT amino acid residue"
FT /label= Unknown
FT /note= "Xaa equals any amino acid residue or no
FT amino acid residue"
FT Misc-difference 217 /label= Unknown
FT /label= Unknown
FT /note= "Xaa equals any amino acid residue or no
FT amino acid residue"
FT
FT
PN WO2000056778-A1.
XX
XX 09-NOV-2000.
XX
XX 27-APR-2000; 2000WO-US11211.
XX
XX 30-APR-1999; 99US-0131966.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J;
XX
XX WPI; 2000-687547/67.
XX
XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
XX for diagnosing, preventing and treating e.g. immune disorders,
FT

PT hyperproliferative disorders and blood coagulation disorders -
XX Claim 11; Fig 2; 277pp; English.
XX
CC The present sequence is a consensus sequence of human, mouse and rat
CC pituitrone. Pituitrone is highly expressed in pituitary gland and is
CC also expressed in brain tissues, spinal cord and kidney. Pituitrone
CC may be used as antigens in the production of antibodies against
CC pituitrone and in assays to identify modulators. Pituitrone cDNAs are
CC also useful in gene therapy. Pituitrone may be used in the prevention,
CC treatment and diagnosis of diseases associated with inappropriate
CC pituitrone expression. It may be useful in treating disorders related
CC to reproductive and renal system, immune disorders, disorders of
CC haematopoietic cells (e.g., anaemia, Digorge syndrome, ataxia
CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation
CC disorders, autoimmune disorders (e.g. Addison's disease, multiple
CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
CC diseases.
XX
SQ Sequence 262 AA;
Query Match 77.0%; Score 1015; DB 21; Length 262;
Best Local Similarity 80.2%; Pred. No. 2e-78;
Matches 210; Conservative 3; Mismatches 47; Indels 2; Gaps 1;
QY 1 MAGSPLLWGPRAGGVGLVLLVLLGLFRPPPALCARPVKPRGLSAASPLAETGAPRRFR 60
DB 1 MAGSPLLXXXXXGCVGLVLLVLLGLXRPXPLXARPVKPRGLSAXSLXETXFXRXR 60
QY 61 RSVPRGEAAGAVQELARALAHLEAEORARAEQAEDQQAQVLAQLLRVWGAPRNSD 120
DB 61 RXVPRGEAAGAVQELARAXAXXLEAEORARAEQAEDQQAQVLAQLLRVWGAPRNSD 120
QY 121 PALGLDDDDPDAPAAQALARALLRDLRALAALQVLPAPVPAALRPPVYDDGAPGPA 180
DB 121 PXLXXDDDDPDAPAAQALARALLRDLRALAALQVLPAPVPAALRPPVYDDGAPGPD 180
QY 181 EAAGDETPDVPPELLRYLLGRILGSLAGSADSEGVAAAP--RLRRAADHDVSGELPPEGLGA 238
DB 181 EXAGDETPDVPPELLRYLLGRILGSLGXXSEPEAAPXXRLRRAVDODLGPVPPENVLGA 240
QY 239 LLRVARLETAPQVPARRLLPP 260
DB 241 LLRVARLENPSPQAPARRLLPP 262
RESULT 8
AAE06088
ID AAE06088 standard; Protein; 148 AA.
XX
AC AAE06088;
XX
DT 24-SEP-2001 (first entry)
XX
DE Human gene 48 encoded secreted protein HKGDL36, SEQ ID NO:150.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vunerary; binding partner identification;
KW gene therapy.
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX Peptide 1..33
XX /label= Signal_peptide

```
FT Protein 34, 148
FT /label= Mature_human_secreted_protein
PN WO200151504-A1.
XX 19-JUL-2001.
XX 12-JAN-2001; 2001WO-US00911.
XX 13-JAN-2000; 2000US-0482273.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX WPI; 2001-425865/45.
DR N-PSDB; AAD11677.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
PT
PS Claim 11; Page 773-774; 864pp; English.
XX
XX AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted
CC protein genes, and AAE06041-AAE06132 represent the proteins they encode.
CC AAE06133-AAE06205 represent human secreted protein fragments.
CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 71 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiocentric disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infectious. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein of
CC the invention.
XX
XX Sequence 148 AA;
XX
XX Query Match 23.4%; Score 308; DB 22; Length 148;
XX Best Local Similarity 50.3%; Pred. No. 1.4e-18;
XX Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;
XX
QY 1 MAGSPLWGRAGVGLVLLGLRPPPPALCARPVKPRGISAASPPLAETG----- 54
DB 1 MAGSPLWGRAGVGLVLLGLRPPPPALCARPVKPRGISAASPPPLARLALLAASG 60
QY 55 --APRRFRSVPV-GBAAGVQELARALAHLEAEQERARAQAEQDQARVLAQLR 111
DB 61 GQCFVRRRCRCRPGAGAGA-----SAGAEQERARAQAR-----LR 98
QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRARLDPAALAAQLVPA 157
DB 99 ISRRASH-----RSCCAGS-----APATLIRLWAWTTPTRLQRSSLAICSAPA 143
```

RESULT 9
ABG33910
ID ABG33910 standard; Protein; 148 AA.
XX
AC ABG33910;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein encode by gene 48.
XX
KW Human; secreted protein; gene therapy; immunosuppressive;
KW antarthritic; antirheumatic; antiproliferative; cytostatic; cardiantic;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; ophthalmological; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US29871.
XX
PR 25-SEP-2000; 2000US-234925P.
PR 12-JAN-2001; 2001WO-US00911.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
DR WPI; 2002-362489/39.
DR N-PSDB; ABK69773.
XX
XX Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT the polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT asthma
XX
PS Claim 11; Page 1252-1253; 1478pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC for example autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. Many other diseases and
CC disorders are listed in the specification. The polypeptides can also be
CC used to aid wound healing an epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative. The polypeptides can increase storage capabilities. The present
CC sequence represents a novel human secreted protein of the invention.
XX
SQ Sequence 148 AA;

Query Match 23.4%; Score 308; DB 23; Length 148;
 Best Local Similarity 50.3%; Pred. No. 1.4e-18;
 Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;

OY 1 MAGSPLLMGPRAGGVLVLLGLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG----- 54
 DB 1 MAGSPLLMGPRAGGVLVLLGLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG 60

OY 55 --APRRFRSVPK-GEAAGAVQELARALAHLLLEAEQERARAEAEDEQOQARVLQAQLLR 111
 DB 61 GOCPEVRRGRRCPCAGAGA-----SAGAERQERARAEQR-----LR 98

OY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRLRDLPAALAAQLVPA 157
 DB 99 ISRRASW-----RSCCAGS-----APPATLIRLWAWTTTPTLRQSSSLALCSAPA 143

Human secreted protein sequence SEQ ID NO:150.

Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;
 tumour; neurodegenerative disorder; developmental abnormality; allergy;
 foetal deficiency; blood disorder; immune system disorder; arthritis;
 autoimmune disease; hepatic disease; renal disease; inflammation;
 Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
 infection; AIDS; spinal cord injury; transplant rejection; diabetes;
 asthma; sepsis; acne; psoriasis; cardiovascular disorder;
 reproductive disorder; gastrointestinal disorder; respiratory disorder;
 metabolic disorder; food additive; preservative.

Homo sapiens.
 WO200004140-A1.
 27-JAN-2000.

14-JUL-1999; 99WO-US15849.
 15-JUL-1998; 98US-0092921.
 15-JUL-1998; 98US-0092922.
 15-JUL-1998; 98US-0092956.
 (HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
 Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
 Mucenski M, Dresser GA, Soppet DR;
 WPI; 2000-161128/14.
 N-PSDB; AAZ98064.

New isolated human genes, useful for diagnosis and treatment of, e.g.
 cancers, neurological or blood disorders -

Claim 11; Page 406-407; 494pp: English.

The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the
 human secreted proteins given in AAZ98064 to AAZ98223. Human secreted
 protein can have activities based on the tissues and cells the genes are
 expressed in. Examples of activities include: cytostatic;
 immunosuppressive; antiinflammatory; nontropic; neuroprotective; and
 antiallergic. The polynucleotides and their corresponding secreted
 polypeptides are useful for preventing, treating or ameliorating medical
 conditions, e.g. by protein or gene therapy. Also pathological conditions

CC can be diagnosed by determining the amount of the new polypeptides in a
 sample or by determining the presence of mutations in the new
 polynucleotides. Human secreted protein s and their polynucleotides can
 be used for developing products for the diagnosis or treatment of cancer,
 tumours, neurodegenerative disorders, developmental abnormalities and
 foetal deficiencies, blood disorders, diseases of the immune system,
 autoimmune diseases, hepatic and renal disease, inflammation,
 allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
 osteoporosis, arthritis, infections, asthma, sepsis, acne, psoriasis,
 transplant rejection, diabetes, reproductive disorders, gastrointestinal
 disorders, respiratory disorders and metabolic disorders. The
 proteins or polynucleotides can also be used as food additives or
 preservatives. The proteins are also useful for identifying their
 binding partners. AAZ98008 to AAZ98016 and AAY87063 are sequence used in
 the exemplification of the present invention.

XX Sequence 149 AA;
 SQ

Query Match 23.4%; Score 308; DB 21; Length 149;
 Best Local Similarity 50.3%; Pred. No. 1.4e-18;
 Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;

OY 1 MAGSPLLMGPRAGGVLVLLGLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG----- 54
 DB 1 MAGSPLLMGPRAGGVLVLLGLLGLFRPPPPALCARPVKEPRGLSAASPPPLAETG 60

OY 55 --APRRFRSVPK-GEAAGAVQELARALAHLLLEAEQERARAEAEDEQOQARVLQAQLLR 111
 DB 61 GOCPEVRRGRRCPCAGAGA-----SAGAERQERARAEQR-----LR 98

OY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRLRDLPAALAAQLVPA 157
 DB 99 ISRRASW-----RSCCAGS-----APPATLIRLWAWTTTPTLRQSSSLALCSAPA 143

Human secreted protein sequence SEQ ID NO:150.

Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;
 tumour; neurodegenerative disorder; developmental abnormality; allergy;
 foetal deficiency; blood disorder; immune system disorder; arthritis;
 autoimmune disease; hepatic disease; renal disease; inflammation;
 Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
 infection; AIDS; spinal cord injury; transplant rejection; diabetes;
 asthma; sepsis; acne; psoriasis; cardiovascular disorder;
 reproductive disorder; gastrointestinal disorder; respiratory disorder;
 metabolic disorder; food additive; preservative.

Homo sapiens.
 WO200004140-A1.
 27-JAN-2000.

14-JUL-1999; 99WO-US15849.
 15-JUL-1998; 98US-0092921.
 15-JUL-1998; 98US-0092922.
 15-JUL-1998; 98US-0092956.
 (HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
 Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
 Mucenski M, Dresser GA, Soppet DR;
 WPI; 2000-161128/14.
 N-PSDB; AAZ98064.

New isolated human genes, useful for diagnosis and treatment of, e.g.
 cancers, neurological or blood disorders -

Claim 11; Page 406-407; 494pp: English.

The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the
 human secreted proteins given in AAZ98064 to AAZ98223. Human secreted
 protein can have activities based on the tissues and cells the genes are
 expressed in. Examples of activities include: cytostatic;
 immunosuppressive; antiinflammatory; nontropic; neuroprotective; and
 antiallergic. The polynucleotides and their corresponding secreted
 polypeptides are useful for preventing, treating or ameliorating medical
 conditions, e.g. by protein or gene therapy. Also pathological conditions

PI Mucenski M, Endress GA, Soppet DR;
 XX WPI; 2000-161128/14.
 DR N-PSDB; AA298064.
 XX
 PT New isolated human genes, useful for diagnosis and treatment of, e.g.
 PT cancers, neurological or blood disorders
 XX
 XX Disclosure; Page 470; 494pp; English.
 PS
 CC The polynucleotide sequences given in AA298017 to AA298108 encode the
 CC human secreted proteins given in AA298064 to AA298723. Human secreted
 CC protein can have activities based on the tissues and cells the genes are
 CC expressed in. Examples of activities include: cytostatic;
 CC immunosuppressive; antiinflammatory; neurotropic; neuroprotective; and
 CC anti-allergic. The polynucleotides and their corresponding secreted
 CC polypeptides are useful for preventing, treating or ameliorating medical
 CC conditions, e.g. by protein or gene therapy. Also pathological conditions
 CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Human secreted proteins and their polynucleotides can
 CC be used for developing products for the diagnosis or treatment of cancer,
 CC tumours, neurodegenerative disorders, developmental abnormalities and
 CC foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC cardiovascular disorders, reproductive disorders, gastrointestinal
 CC disorders, respiratory disorders and metabolic disorders. The
 CC proteins or polynucleotides can also be used as food additives or
 CC preservatives. The proteins are also useful for identifying their
 CC binding partners. AA298008 to AA298016 and AA298063 are sequence used in
 CC the exemplification of the present invention.
 XX
 XX Sequence 166 AA;
 SQ
 Query Match 23.4%; Score 308; DB 21; Length 166;
 Best Local Similarity 50.3%; Pred. No. 1.6e-18;
 Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;
 QY 1 MAGSPLWGPAGVGLVLLVLLGLFRPPALCARPVKEPRGLSAASPPLAETG----- 54
 DB 19 MAGSPLWGPAGVGLVLLVLLGLFRPPALCARPVKEPRGLSAASPPLARLALLAASG 78
 QY 55 --APRRFRSVP-GEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLLR 111
 DB 79 GQCPVEVRRRCRPGAGAGA-----SAGAERQERARAQAQ-----LR 116
 QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR----ALLRARLDPAALAAQLVPA 157
 DB 117 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTTRLQRSSSLALCSAPA 161
 RESULT 12
 AA2980176
 ID AAE06176 standard; Protein; 166 AA.
 XX
 AC AAE06176;
 XX
 XX 24-SEP-2001 (first entry)
 DT
 XX Human gene 48 encoded secreted protein fragment, SEQ ID NO:238.
 DE
 XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;
 KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KW cell culture; chemotaxis; vulnery; binding partner identification;

KW gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200151504-A1.
 XX
 XX 19-JUL-2001.
 PD
 XX
 PF 12-JAN-2001; 2001WO-US00911.
 XX
 PR 13-JAN-2000; 2000US-0482273.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;
 PI Lafleur DW, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;
 PI Endress GA, Mucenski M, Ebner R;
 XX
 XX WPI; 2001-425865/45.
 DR
 XX Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 PT
 PS Disclosure; Page 122; 864pp; English.
 XX
 CC AAD11630-AAD11721 represent cDNAs corresponding to 71 human secreted
 CC protein genes, and AAE06041-AAE06132 represent the proteins they encode.
 CC AAE06133-AAE06205 represent human secreted protein fragments.
 CC The secreted proteins and their genes are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 71 genes,
 CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders, cancer, tumours, foetal and developmental abnormalities,
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
 CC The present sequence represents a human secreted protein fragment
 CC referred to in the disclosure of the invention.
 XX
 SQ Sequence 166 AA;
 Query Match 23.4%; Score 308; DB 22; Length 166;
 Best Local Similarity 50.3%; Pred. No. 1.6e-18;
 Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;
 QY 1 MAGSPLWGPAGVGLVLLVLLGLFRPPALCARPVKEPRGLSAASPPLAETG----- 54
 DB 19 MAGSPLWGPAGVGLVLLVLLGLFRPPALCARPVKEPRGLSAASPPLARLALLAASG 78
 QY 55 --APRRFRSVP-GEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLLR 111
 DB 79 GQCPVEVRRRCRPGAGAGA-----SAGAERQERARAQAQ-----LR 116
 QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR----ALLRARLDPAALAAQLVPA 157
 DB 117 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTTRLQRSSSLALCSAPA 161

RESULT 13
ABG34000
ID ABG34000 standard; Protein; 166 AA.
XX
XX
AC ABG34000;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein #2 encoded by gene 48.
XX
XX Human; secreted protein; gene therapy; immunosuppressive;
KW antichronic; antineumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; ophthalmological; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
XX
PD 04-APR-2002.
XX
XX 24-SEP-2001; 2001WO-US29871.
XX
XX 25-SEP-2000; 2000US-234925P.
PR 12-JAN-2001; 2001WO-US00911.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
XX WPI; 2002-362489/39.
DR N-PSDB; ABK69773.
XX
XX Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT the polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT asthma
XX
XX Disclosure; Page 1316-1317; 1478pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC for example autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. Many other diseases and
CC disorders are listed in the specification. The polypeptides can also be
CC used to aid wound healing an epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence represents a novel human secreted protein of the invention.
XX
XX Sequence 166 AA;
SQ

Query Match 23.4%; Score 308; DB 23; Length 166;
Best Local Similarity 50.3%; Pred. No. 1.6e-18;
Matches 88; Conservative 3; Mismatches 34; Indels 50; Gaps 8;
QY 1 MAGSPLLMGPRAGGVLVLLVLLGLFRPPPPALCARPVKEPGLSASPPLAETG----- 54
DB 19 MAGSPLLMGPRAGGVLVLLVLLGLFRPPPPALCARPVKEPGLSASPPPLARLALIASG 78
QY 55 --APRRERSVPR-GEAGAGVQELARALAHLEAEPRQERARAEAEQDQARVLAQLLR 111
DB 79 GOCPEVRRRCRCPGAGAGA-----SAGAEQERARAEQOR-----LR 116
QY 112 V-----WGAPRNSDPALGLDDDDPAPAAQLAR-----ALLRLRDLPAALAAQLVPA 157
DB 117 ISRRASW-----RSCCAGS-----APPATLRLWAWTTTPTLRLQSSSLALCSAPA 161
RESULT 14
AAV71961
ID AAV71961 standard; Protein; 69 AA.
XX
XX
AC AAV71961;
XX
DT 26-MAR-2001 (first entry)
XX
DE Rat pituitary hormone, pituitrone.
XX
XX Rat; pituitary; pituitrone; therapy; immune disorder; anaemia;
KW Digeeorge syndrome; haematopoietic cell; ataxia telangiectasia;
KW Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW SLE; hyperproliferative disorder; gene therapy; neoplasm;
KW infectious disease; immunomodulatory; cytostatic; antimicrobial.
XX
OS Rattus sp.
XX
XX WO2000066778-A1.
PN
XX
PD 09-NOV-2000.
XX
XX 27-APR-2000; 2000WO-US11211.
PF
XX
XX 30-APR-1999; 99US-0131966.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Ni J;
PI
XX
XX WPI; 2000-687547/67.
DR N-PSDB; AAD02077.
XX
XX Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
PT for diagnosing, preventing and treating e.g. immune disorders,
PT hyperproliferative disorders and blood coagulation disorders -
XX
XX Claim 11; Fig 2; 277pp; English.
XX
XX The present sequence is a rat pituitary hormone, pituitrone.
CC pituitrone is highly expressed in pituitary gland and is
CC also expressed in brain tissues, spinal cord and kidney. It may be
CC used as antigens in the production of antibodies against pituitrone
CC and in assays to identify modulators. Pituitrone cDNAs are
CC also useful in gene therapy. Pituitrone may be used in the prevention,
CC treatment and diagnosis of diseases associated with inappropriate
CC pituitrone expression. It may be useful in treating disorders related
CC to reproductive and renal system, immune disorders, disorders of
CC haematopoietic cells (e.g., anaemia, Digeeorge syndrome, ataxia
CC telangiectasia and Wiskott-Aldrich disorder), blood coagulation
CC disorders, autoimmune disorders (e.g. Addison's disease, multiple
CC sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
CC disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
CC diseases.

XX Sequence 69 AA;
S0
Query Match 19.6%; Score 259; DB 21; Length 69;
Best Local Similarity 75.0%; Pred. No. 8.2e-15;
Matches 51; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
QY 193 ELLRYLLGRILGASDSEGVAAAPRRRLRAADHDVSGELPPEGVLGALLRVKRLTETAPQV 252
DB 2 ELLRYLLGRILGASDSEGVAAAPRRRLRAADHDVSGELPPEGVLGALLRVKRLTETAPQV 61
QY 253 PARRLLPP 260
DB 62 PARRLLPP 69
RESULT 15
ABG60300
ID ABG60300 standard; Protein; 388 AA.
XX
AC ABG60300;
XX
DT 30-JUL-2002 (first entry)
XX
DE Lymphona associated human G protein XI alpha S.
XX
KW Lymphona associated protein; Pik3r1; GNAS; XI(alpha_s.); NESP55; JAK1;
KW Neurogranin; Nfr2; lymphoma; DNA vaccine; cytotoxic T-cell;
KW animal model; Leukaemia.
XX
OS Homo sapiens.
XX
PN WO200224867-A2.
XX
PD 28-MAR-2002.
XX
PF 24-SEP-2001; 2001WO-US29798.
XX
PR 22-SEP-2000; 2000US-0668644.
PR 13-JUL-2001; 2001US-0905390.
PR 13-JUL-2001; 2001US-0905491.
XX
PA (UYAA-) UNIV AARHUS.
XX
PI Pedersen FS, Sorensen AB, Hernandez JM, Nielsen AA, Moving HO;
XX
WPI; 2002-416279/44.
DR N-PSDB; ABR72323.
XX
PT Novel recombinant lymphoma associated protein (LAP) such as Pik3r1,
PT GNAS, JAK1, Neurogranin, Nfr2 proteins, useful for identifying
PT inhibitors of LAP activity that are used for treating lymphoma -
XX
PS Claim 5; Page 136; 160pp; English.
XX
CC The invention describes a recombinant protein (I) from a lymphoma
CC associated protein (LAP) sequence of Pik3r1, GNAS (including
CC XI(alpha_s.), and NESP55), JAK1, Neurogranin or Nfr2 proteins. (I) is
CC useful for screening for a bioactive agent capable of binding to an LA
CC protein (LAP) which is encoded by a polynucleotide (II) and (II) is
CC useful for evaluating the effect of a candidate lymphoma drug in a
CC patient. (I) and (II) is also useful for diagnosing lymphoma involving
CC determining the expression of one or more (II), or (I) encoded by (II)
CC in a first tissue type of a first individual and comparing the expression
CC of the gene(s) from a second normal tissue type from the first individual
CC or a second unaffected individual, where a difference in the expression
CC indicates that the first individual has lymphoma. (II) is useful in the
CC diagnosis and treatment of lymphoma and leukaemia. A host cell
CC comprising (II) is useful for screening drug candidates which involves
CC providing the host cell that expresses LA gene, adding a drug candidate
CC to the cell, and determining the effect of the drug candidate on the
CC expression of the LA gene. LA genes are also administered as DNA
CC vaccines, such that expression of the polypeptide encoded by the DNA

CC vaccine, cytotoxic T-cells and antibodies are induced which recognise and
CC destroy or eliminate cells expressing LA proteins. The genes are also for
CC generating animal models of lymphoma. This sequence represents a lymphona
CC associated protein described in the invention.
XX
SQ Sequence 388 AA;
Query Match 10.3%; Score 136.5; DB 23; Length 388;
Best Local Similarity 27.5%; Pred. No. 0.0016;
Matches 67; Conservative 19; Mismatches 89; Indels 69; Gaps 10;
QY 47 SPPLAETGAPRRFRSVPRCGACAGVQELARALAHLLAEERQGRARAEQAEEDQQAARVL 106
DB 26 SPPIALDGPPIK-----VSGAPD-----KREARPPVEEAAEMEGAADA 66
QY 107 AQLLRV-----WGAP-----RNSDPALG-LDDDDPDAPAAQLARALLRRL 145
DB 67 AEGGKVPSPGCGSPAAGASADTAARAAPADPDGSGATPEDPDGCTAP-----ADP 119
QY 146 DPAALAAQLVPAPVPAALRPVDDGDPAGDPAEAGDETPDVPPELLRYLLGRILAG 205
DB 120 DSGAFAADPDGGAAPADPDGGAAPADPDGGAAPADPDGGAAPADPD-----AG 167
QY 206 SADSEGVAAAPRRRLRAADH-----DVGSELPEGVGALLRVKRLTETAPQVPARR--- 256
DB 168 AA-PEAPAAPAAAEETRAAHVAPAAPDAGAPTAAASATRAAQVNRRAASAPASGARRKH 226
QY 257 LLPP 260
DB 227 LRPP 230

Search completed: April 4, 2003, 14:16:20
Job time : 39 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 18:34:13 ; Search time 274 Seconds
(without alignments)
8046.374 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcggcagcagggccagc.....gaggatccctaccctctggc 979

Scoring table:

Gapop 60.0 , Capext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 15

Total number of hits satisfying chosen parameters: 8227

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

```

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	979	100.0	979	20 AAV68668	Nucleotide sequenc
2	891	91.0	1030	22 AAI58161	Human polynucleoti
3	891	91.0	1041	22 AAI59947	Human polynucleoti
4	891	91.0	1041	22 AAI59948	Human polynucleoti
5	891	91.0	1043	21 AAD02075	cDNA encoding huma
6	840	85.8	998	22 AAF94496	Human hydrophobic
7	678	69.3	780	22 AAF94486	Human hydrophobic
8	583	59.6	1050	21 AAZ98064	Human secreted pro
9	583	59.6	1050	22 AAD11677	Human secreted pro

10	583	59.6	1050	24	ABK69773	Human secreted pro
11	588	58.0	919	22	AAI58162	Human polynucleoti
12	375	38.3	4503	24	ABK69876	Human secreted pro
13	226	23.1	301	24	ABN95203	Gene #1701 used to
14	226	23.1	301	24	ABL68387	Kidney cancer rela
15	60	6.1	60	24	ABN42141	Human spliced tran
16	32	3.3	991	21	AAD02076	cDNA encoding mous
17	24	2.5	652	24	ABQ31156	Oligonucleotide fo
18	24	2.5	652	24	ABQ31157	Oligonucleotide fo
19	23	2.3	652	24	ABQ31154	Oligonucleotide fo
20	23	2.3	652	24	ABQ31155	Oligonucleotide fo
21	22	2.2	425	22	AAF65901	Novel human polynu
22	22	2.2	1507	20	AAH84446	Mouse brain CNG-3
23	21	2.1	422	18	AAH88808	Nuclear steroid ho
24	21	2.1	533	22	AAD03545	Human EST #3 to pr
25	20	2.0	355	22	AAF65178	Novel human polynu
26	20	2.0	396	21	AAD02077	cDNA encoding rat
27	20	2.0	742	20	AAZ15143	Human gene express
28	19	1.9	156	23	AAH5491	DNA encoding novel
29	19	1.9	156	23	AAH5491	DNA encoding novel
30	19	1.9	156	24	AAH63043	Cell death protect
31	19	1.9	251	19	AAH10968	Human biallelic po
32	19	1.9	255	20	AAH87627	Human single nucle
33	19	1.9	300	20	AAZ15019	Human gene express
34	19	1.9	349	24	AAH61708	Lung small cell ca
35	19	1.9	380	24	ABK62892	Rat sequence diffe
36	19	1.9	383	22	AAF64866	Novel human polynu
37	19	1.9	389	22	AAF64961	Novel human polynu
38	19	1.9	398	22	AAF64559	Novel human polynu
39	19	1.9	425	22	AAF64749	Novel human polynu
40	19	1.9	463	23	AAH5492	DNA encoding novel
41	19	1.9	464	24	AAH61489	Lung small cell ca
42	19	1.9	467	20	AAV64423	Mouse developing l
43	19	1.9	476	24	ABN95942	Gene #2440 used to
44	19	1.9	486	19	AAV40148	DNA sequence (crtE
45	19	1.9	489	16	AAQ99486	3 hydroxy-beta-ion

ALIGNMENTS

RESULT 1

AAV68668
ID AAV68668 standard; DNA; 979 BP.

AC AAV68668;

DT 03-JUN-1999 (first entry)

DE Nucleotide sequence of the human Tango-81 gene.

KW Human; Tango-81; host cell; recombinant protein; antibody;

KW receptor; specific binding agent; probe; primer; hybridisation;

KW amplification; mutation; genetic mapping; ss.

OS Homo sapiens.

FX Key Location/Qualifiers

FT CDS 58..840

FT /*tag- a

FT /product= "Tango-81"

PN WO9906427-A1.

XX 11-FEB-1999.

XX 04-AUG-1998; 98WO-US16241.

XX 04-AUG-1997; 97US-0054645.

XX (MILL-) MILLENNIUM BIOTHEAPEUTICS INC.

XX McCarthy SA;

PI

PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 364; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1030 BP; 134 A; 404 C; 352 G; 140 T; 0 other;
Query Match 91.0%; Score 891; DB 22; Length 1030;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 38 GCCCGGCTCGCTGGGAGCAGATCGCGGGTCGCGGCTGCTGGGGCCGCGGGCGGGG 97
DB 53 GCCCGGCTCGCTGGGAGCAGATCGCGGGTCGCGGCTGCTGGGGCCGCGGGCGGGG 112
QY 98 GCGTCGGCCTTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCG 157
DB 113 GCGTCGGCCTTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCG 172
QY 158 CGCGCGCGGTAAAGGAGCCCGCGGCTCAGCGACGCTCTCCGCCCTTGGCTGACACTG 217
DB 173 CGCGCGCGGTAAAGGAGCCCGCGGCTCAGCGACGCTCTCCGCCCTTGGCTGACACTG 232
QY 218 GCGCTCTCGCCCTTCCGCGGCTCAGTCCCGAGGTGAGCGCGGGGCGCGCTGCAGG 277
DB 233 GCGCTCTCGCCCTTCCGCGGCTCAGTCCCGAGGTGAGCGCGGGGCGCGCTGCAGG 292
QY 278 AGCTGGCGGGGCGCTGGGCGCATCTGCTGGAGCCGAAAGTCAGAGCGGGGCGGGGCGG 337
DB 293 AGCTGGCGGGGCGCTGGGCGCATCTGCTGGAGCCGAAAGTCAGAGCGGGGCGGGGCGG 352
QY 338 AGCGCAGGAGGCTGAGGATCAGCAGCGCGCTCTGCGCAGCTGCTGCGCGCTGCGG 397
DB 353 AGCGCAGGAGGCTGAGGATCAGCAGCGCGCTCTGCGCAGCTGCTGCGCGCTGCGG 412
QY 398 GCGCCCCCGCAACTCTGATCCGGCTCTGGGCTTGGACGACGACCCCGAGCGCGCTGCAG 457
DB 413 GCGCCCCCGCAACTCTGATCCGGCTCTGGGCTTGGACGACGACCCCGAGCGCGCTGCAG 472
QY 458 CGCAGCTCGCTCGCTCTGCTCGCGCGCGCTTGAACCTGCGCGCTAGCAGCCGAGC 517
DB 473 CGCAGCTCGCTCGCTCTGCTCGCGCGCGCTTGAACCTGCGCGCTAGCAGCCGAGC 532
QY 518 TTGTCCCGCGCGCGCTCCCGCGCGCGCTCCGACCCCGCGCGCTACGACGACG 577
DB 533 TTGTCCCGCGCGCGCTCCCGCGCGCGCTCCGACCCCGCGCGCTACGACGACG 592
QY 578 GCCCGCGGGCCCGGATGCTGAGGAGCGAGCGGACGACACCCGAGCTGCAGCCCGAGC 637
DB 593 GCCCGCGGGCCCGGATGCTGAGGAGCGAGCGGACGACACCCGAGCTGCAGCCCGAGC 652
QY 638 TGTGTAGGTACTTGTGAGGACGAGTCTTTCGGGGAAGCGCGGACTCCGAGGGGGTGGCAG 697
DB 653 TGTGTAGGTACTTGTGAGGACGAGTCTTTCGGGGAAGCGCGGACTCCGAGGGGGTGGCAG 712
QY 698 CCGCGCGCGCGCTCCCGCGCGCGGACCATGCTGGCTGAGCTGCCCGCTGAGG 757
DB 712 CCGCGCGCGCGCTCCCGCGCGCGGACCATGCTGGCTGAGCTGCCCGCTGAGG 757

DB 713 CCGCGCGCGCGCTCCCGCGCGCGCTCCGACACGATGTGGGCTCTGAGCTGCCCGCTGAGG 772
QY 758 GCGTCTGGGGGCGCTGCTGCTGTGTAAGCGCTAGAGACCCCGGGCCCGCAGGTGCTG 817
DB 773 GCGTCTGGGGGCGCTGCTGCTGTGTAAGCGCTAGAGACCCCGGGCCCGCAGGTGCTG 832
QY 818 CAGCGCGCTCTTGGCCACCTGAGCACTGCGCGGATCCCGTGCACCTTGGGACCCAGAAG 877
DB 833 CAGCGCGCTCTTGGCCACCTGAGCACTGCGCGGATCCCGTGCACCTTGGGACCCAGAAG 892
QY 878 TGCGCCCGCGCATCCCGCCACCGAGCTGCTCCCGCGCAGCAGCTCCAGAGCAACTTACCC 937
DB 893 TGCGCCCGCGCATCCCGCCACCGAGCTGCTCCCGCGCAGCAGCTCCAGAGCAACTTACCC 952
QY 938 CGCGCAGCGCAGCGCTCTCACCAGGATCCCTACCCCGCTGGC 979
DB 953 CGCGCAGCGCAGCGCTCTCACCAGGATCCCTACCCCGCTGGC 994
RESULT 3
AAI59947/c
ID AAI59947 standard; cDNA; 1041 BP.
XX
AC AAI59947;
XX
DT 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 3936.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM40791.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 3936; 10078pp; English.
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and

XX	Sequence 1041 BP; 145 A; 356 C; 402 G; 138 T; 0 other;	
SY	Query Match 91.0%; Score 891; DB 22; Length 1041;	
	Best Local Similarity 99.9%; Pred. No. 0;	
	Matches 941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	38 GCGCGCTCGTGGGACAGATGGGGTTCGGCTGCTGCGGGGCGCGGGCGGGG 97	
DB	981 GCGCGCTCGTGGGACAGATGGGGTTCGGCTGCTGCGGGGCGCGGGCGGGG 922	
QY	98 GCGTGGCTTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 157	
DB	921 GCGTGGCTTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 862	
QY	158 GCGCGCGGTAAAGGAGCGCGCGCGCTAAGCGACAGCTCCGCCCTTGGCTGAGACTG 217	
DB	861 GCGCGCGGTAAAGGAGCGCGCGCGCTAAGCGACAGCTCCGCCCTTGGCTGAGACTG 802	
QY	218 GCGCTCTCGCGCTTCGGCGGTTCAGTCCCGGAGGTGAGCGGGGCGGGGCGGTCAGG 277	
DB	801 GCGCTCTCGCGCTTCGGCGGTTCAGTCCCGGAGGTGAGCGGGGCGGGGCGGTCAGG 742	
QY	278 AGCTGCGCGGGCGTGGCGCATCTGCTGAGCGCAAGCTCAGGAGCGGCGCGGCGG 337	
DB	741 AGCTGCGCGGGCGTGGCGCATCTGCTGAGCGCAAGCTCAGGAGCGGCGCGGCGG 682	
QY	338 AGCGCAGGAGGCTGAGGATCAGCAGCGCGGCTCCTGGCGAGCTGCTGCGCGTCTGGG 397	
DB	681 AGCGCAGGAGGCTGAGGATCAGCAGCGCGGCTCCTGGCGAGCTGCTGCGCGTCTGGG 622	
QY	398 GCGCGCGCGCGGCTGATCGCGCTTGGGCTTGGACAGCAGCAGCGCGCGCTGCGAG 457	
DB	621 GCGCGCGCGCGGCTGATCGCGCTTGGGCTTGGACAGCAGCAGCGCGCGCTGCGAG 562	
QY	458 GCGAGCTCGCTGCGCTGCTCGCGCGCGCTTGCACCTGCGCGCTAGCAGCGCCAGC 517	
DB	561 GCGAGCTCGCTGCGCTGCTCGCGCGCGCTTGCACCTGCGCGCTAGCAGCGCCAGC 502	
QY	518 TTGTCCCGCGCGCTGCTCGCGCGCGCTCGGACCCCGCGCGCTTACGAGGAGG 577	
DB	501 TTGTCCCGCGCGCTGCTCGCGCGCGCTCGGACCCCGCGCGCTTACGAGGAGG 442	
QY	578 GCGCGCGCGCGGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 637	
DB	441 GCGCGCGCGCGGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 382	
QY	638 TGTGAGTACTTGTGCGGACGATTTTTCGGGAGCGCGGACTCCGAGGGGGTGGCAG 697	
DB	381 TGTGAGTACTTGTGCGGACGATTTTTCGGGAGCGCGGACTCCGAGGGGGTGGCAG 322	
QY	698 GCGCGCGCGCTTCCGCGCTGCGCGGACGACGATGTGGGCTGTGAGCTGCCCGCTGAGG 757	
DB	321 GCGCGCGCGCTTCCGCGCTGCGCGGACGACGATGTGGGCTGTGAGCTGCCCGCTGAGG 262	
QY	758 GGTGCTGGGGGCGCTGCTGCTGTAACGCTAGACACCCCGCGCGCGCGCGCTGCTG 817	
DB	261 GGTGCTGGGGGCGCTGCTGCTGTAACGCTAGACACCCCGCGCGCGCGCGCTGCTG 202	
QY	818 CAGCGCGCGCTTTCGACCGCTGAGCAGCTCCGCGGATCCCGCTGACCGCTGGGACCCAGAGA 877	
DB	201 CAGCGCGCGCTTTCGACCGCTGAGCAGCTCCGCGGATCCCGCTGACCGCTGGGACCCAGAGA 142	
QY	878 TGCCTCCCGCGCTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 937	
DB	141 TGCCTCCCGCGCTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 82	
QY	938 GCGCGAGCGCGCTTTCACCGGAGGATCCCTACCGCTGGC 979	
DB	81 GCGCGAGCGCGCTTTCACCGGAGGATCCCTACCGCTGGC 40	

RESULT 5

AA02075	standard; cDNA; 1043 BP.
XX	AA02075;
XX	26-MAR-2001 (first entry)
DE	cDNA encoding human pituitary hormone, pituitrone.
XX	Human; pituitary; pituitrone; therapy; immune disorder; anaemia;
KW	Digeorge syndrome; haematopoietic cell; ataxia telangiectasia;
KW	Wiskott-Aldrich disorder; blood coagulation disorder; Addison's disease;
KW	autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW	SLE; hyperproliferative disorder; gene therapy; neoplasm;
KW	infectious disease; immunomodulatory; cytostatic; antimicrobial; ss.
OS	Homo sapiens.
XX	
PH	Key Location/Qualifiers
FT	CDS 44..826
FT	/tag= a
FT	/product= "Human pituitary hormone, pituitrone"
FT	44..142
FT	/tag= b
FT	143..823
FT	/tag= c
FT	/product= "Mature human pituitary hormone, pituitrone"
XX	W0200066778-A1.
PN	09-NOV-2000.
XX	
PF	27-APR-2000; 2000WO-US11211.
XX	
PR	30-APR-1999; 99US-0131966.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Ruben SM, Ni J;
XX	
DR	WPI; 2000-687547/67.
DR	P-PSDB; AAY1959.
XX	
PT	Nucleic acids encoding a novel pituitary hormone (pituitrone), useful
PT	for diagnosing, preventing and treating e.g. immune disorders,
PT	hyperproliferative disorders and blood coagulation disorders -
XX	
PS	Claim 4; Fig 1; 277pp; English.
XX	
CC	The present sequence is a cDNA clone HKDL36 encoding human pituitary
CC	hormone, pituitrone. Pituitrone is highly expressed in pituitary
CC	gland and is also expressed in brain tissues, spinal cord and kidney.
CC	It may be used as antigens in the production of antibodies against
CC	pituitrone and in assays to identify modulators. Pituitrone cDNAs are
CC	also useful in gene therapy. Pituitrone may be used in the prevention,
CC	treatment and diagnosis of diseases associated with inappropriate
CC	pituitrone expression. It may be useful in treating disorders related
CC	to reproductive and renal system, immune disorders, disorders of
CC	haematopoietic cells (e.g., anaemia, Digeorge syndrome, ataxia
CC	telangiectasia and Wiskott-Aldrich disorder), blood coagulation
CC	disorders, autoimmune disorders (e.g. Addison's disease, multiple
CC	sclerosis and systemic lupus erythematosus (SLE)), hyperproliferative
CC	disorders (e.g. neoplasms of bone, liver and pancreas) and infectious
CC	diseases.
XX	
SY	Sequence 1043 BP; 171 A; 391 C; 346 G; 135 T; 0 other;
	Query Match 91.0%; Score 891; DB 21; Length 1043;
	Best Local Similarity 99.9%; Pred. No. 0;
	Matches 941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	38 GCGCGCTCGTGGGACAGATGGGGTTCGGCTGCTGCGGGGCGCGGGCGGGG 97

Db	24	CCCCGCTCGCTGGGCGAGCATGGCGGGTGTGCCGCTGTCTTGGGGGGCGCGCGGGCGGGGCG 83
Qy	98	GCFTCGGCCTTTTGGTGTCTGTGTCTCGGCTGTCTTGGCGCGCGCCCGCGGCTCTGCG 157
Db	84	GCFTCGGCCTTTTGGTGTCTGTGTCTCGGCTGTCTTGGCGCGCCCGCGGCTCTGCG 143
Qy	158	CGCGCGGTAAAGGAGCCCGCGGCTTAAGCGACGGTCTCCGCGCTTGGCTGAGACTG 217
Db	144	CGCGCGGTAAAGGAGCCCGCGGCTTAAGCGACGCTCTCGCGCTTGGCTGAGACTG 203
Qy	218	GCCTCTCTCGCGCTTCCGGGCTCAGTGCCTCGAGGTGAGGCGGCGGGGCGGTGCAG 277
Db	204	GCCTCTCTCGCGCTTCCGGGCTCAGTGCCTCGAGGTGAGGCGGCGGGGCGGTGCAG 263
Qy	278	AGCTGGCGGGGCGCTGGCGCATCTGCTGAGGCGCAAGCTCAGGAGCGGCGCGGGCGG 337
Db	264	AGCTGGCGGGGCGCTGGCGCATCTGCTGGAGGCGCAAGCTCAGGAGCGGCGGGGCGG 323
Qy	338	AGGCGCAGGAGGTGAGGATCAGCAGCGCGCTCTTGCGCAGCTGCTGCCTGTGGG 397
Db	324	AGGCGCAGGAGGTGAGGATCAGCAGCGCGCTCTTGCGCAGCTGCTGCCTGTGGG 383
Qy	398	GCGCCCGCCGCACTCTGATCCGGCTTGGGCTTGGACAGCAGACCCGAGCGGCTGCAG 457
Db	384	GCGCCCGCCGCACTCTGATCCGGCTTGGGCTTGGACAGCAGACCCGAGCGGCTGCAG 443
Qy	458	CGCAGCTCGCTCGCGCTCTGCTCCGCGCGCGCTTGCACCTTGCGGCGCTAGCAGCCACG 517
Db	444	CGCAGCTCGCTCGCGCTCTGCTCCGCGCGCGCTTGCACCTTGCGGCGCTAGCAGCCACG 503
Qy	518	TTGTTCCTCCGCGCGCTCCCGCGCGCGGCTCGACCCGGCCCCGGTCTACGACGAG 577
Db	504	TTGTTCCTCCGCGCGCTCCCGCGCGCGGCTCGACCCGGCCCCGGTCTACGACGAG 563
Qy	578	GCCCGGGGCGCGATGCTGAGGAGCGGACAGACACCCGACGTGGACCCCGGAGC 637
Db	564	GCCCGGGGCGCGATGCTGAGGAGCGGACAGACACCCGACGTGGACCCCGGAGC 623
Qy	638	TGTTGAGGTACTTGTCTGGGACGGATTCTTGGGGAAGCGGACTCCGAGGGGTGGAC 697
Db	624	TGTTGAGGTACTTGTCTGGGACGGATTCTTGGGGAAGCGGACTCCGAGGGGTGGAC 683
Qy	698	CCCGCGCGCGCTCCGCGGTGCGCGACACAGATGCGGCTTGTGCTGCCCCCTGAGG 757
Db	684	CCCGCGCGCGCTCCGCGGTGCGCGACACAGATGCGGCTTGTGCTGCCCCCTGAGG 743
Qy	758	GCCTGTGGGGGCGCTGCTGTGTGAACGCTTAGAGACCCCGGGGCGCCAGGTGCCTG 817
Db	744	GCCTGTGGGGGCGCTGCTGTGTGAACGCTTAGAGACCCCGGGGCGCCAGGTGCCTG 803
Qy	818	CAGCGCGCTTGTCCACCTGAGCACTGCCGGATCCGCTGCACCTTGGGACCCAGAG 877
Db	804	CAGCGCGCTTGTCCACCTGAGCACTGCCGGATCCGCTGCACCTTGGGACCCAGAG 863
Qy	878	TGCCCCCGGCATCCCGCCACGAGCTGTCTCCCGCGCAGCAGCTCCAGAGCAACTTACC 937
Db	864	TGCCCCCGGCATCCCGCCACGAGCTGTCTCCCGCCAGCAGCTCCAGAGCAACTTACC 923
Qy	938	CGCGCAGCAGCCCTCTCACCGGAGATCCCTTACCCCTTGGC 979
Db	924	CGGCCAGCAGCCCTCTCACCGGAGATCCCTTACCCCTTGGC 965

RESULT 6

AAAF94496

ID AAF94496 standard; cDNA; 998 BP.

AA
AC AAF94496;

DT 04-JUN-2001 (first entry)

DE Human hydrophobic domain containing protein clone HP10756 cDNA #120.

XX

www

•

•

Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective; antianemic; vulnery; anticler; osteopathic; anti-inflammatory; cystostatic; gene therapy; autoimmune disorder; multiple sclerosis; HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing; inflammatory bowel disease; nutritional supplement; appetite; vaccine; behavioural characteristic; immune response; ss.

Homo sapiens.

WO200112660-A2.

22-FEB-2001.

10-AUG-2000; 2000WO-JP05356.

17-AUG-1999; 99JP-0230344.

17 AUG 1999; 99JP-0230344;
07-SEP-1999; 99JP-0252551.

01-OCT-1999; 99JP-0281132.

22-OCT-1999; 99JP-0301624.
04-NOV-1999. 99JP-0313877

04-NOV-1999; 99JP-0313877.

(SAGA) SAGAMI CHEM RES CENT.
(PROT-) PROTEGENE INC.

Kato S, Kimura T;

WPI; 2001-160059/16.

P-PSDB; AAB88596.

Human proteins with hydrophobic domains and the DNAs which encode them are useful for treating autoimmune disorders, burns and tumors and for screening novel pharmaceuticals -

Claim 4; Page 461-463; 518pp; English.

AA9794417 to AAF94516 encode the human proteins given in AAB88557 to AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant, anti-HIV, neuroprotective, antianemic, vulnery, antitumor, antileukemic, osteoplastic, anti-inflammatory and cytostatic activities, and can be used in gene therapy. (I) can be used as pharmaceuticals and as antigens to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as probes for genetic diagnosis and gene sources for gene therapy or for producing (I) in large quantities. Cells containing (II) are used for the detection of ligands or receptors corresponding to membrane or secretory proteins and to screen small molecule novel pharmaceuticals.

Sequence 998 BP; 120 A; 394 C; 347 G; 137 T; 0 other;

Query Match 85.8%; Score 840; DB 22; Length 998;

Query Match: 99.8%; Score: 840;
Best Local Similarity 99.8%; Pred. No. 0;

100% local similarity	99.96%	freq. NO: 0;
Mismatches	940;	Conservative
Mismatches	0;	Mismatches
Indels	2;	Indels
Gaps	0;	Gaps

AC AAF94496;

Y 50

DT 04-JUN-2001 (first entry)

30 G C C C G G C T C G C T G G G G C A G C A T G G C G G G G T C G C C G C T G C T C T G G G G G C C G C G G G G G

98 GCGTCGGCCTTTGGTGCTGCTGCTCGGCCCTGTTTCGGCGCCCCCGCGCTCTGCGG

Db 121 CGCGCCCTAAGCAGCGCTCTCCGCCCTTGCTGAGACTGGCGCTCTCTCGCGCTTCGGG 180

Qy 238 CGGTAGTGGCCCCGAGGTAGCGCGCGGGCGGTGAGAGCTGGCGCGCGCTGGCG 297

Db 181 CGGTAGTGGCCCCGAGGTAGCGCGCGGGCGGTGAGAGCTGGCGCGCGCTGGCG 240

Qy 298 CATCTGCTGGAGGCGGACGTCAGAGGCGGCGGGCGGCGGAGGCGAGGAGCTGAGGAT 357

Db 241 CATCTGCTGGAGGCGGACGTCAGAGGCGGCGGGCGGCGGAGGCGAGGAGCTGAGGAT 300

Qy 358 CAGCAGCGCGCGCTTCGCGCGCTGCTGCGCTGCTGGCGCGCGCGCGCAACTCTGAT 417

Db 301 CAGCAGCGCGCGCTTCGCGCGCTGCTGCGCTGCTGGCGCGCGCGCGCAACTCTGAT 360

Qy 418 CCGGCTCTGGGCTTGGAGCAGACCGCGCGCTGCGAGCGAGCTGCTGCGCGCTG 477

Db 361 CCGGCTCTGGGCTTGGAGCAGACCGCGCGCTGCGAGCGAGCTGCTGCGCGCTG 420

Qy 478 CTCCGGCGCGCTTGACCTTGGCGCGCTAGCAGCGCGCTGCTGCGCGCGCGCGTCCC 537

Db 421 CTCCGGCGCGCGCTTGACCTTGGCGCGCTGCGAGCGCGCGCGCGCGCGCGTCCC 480

Qy 538 GCCCGCGCGCTTCCGCGCGCGCGCTTACGACGAGCGCGCGCGCGCGCGGATGCT 597

Db 481 GCCCGCGCGCTTCCGCGCGCGCGCTTACGACGAGCGCGCGCGCGCGCGGATGCT 540

Qy 598 GAGGAGCGCGCGAGCAGACCGCGCGCTGCGAGCGCGCGCGCGCGCTTCTGCTGGA 657

Db 541 GAGGAGCGCGCGAGCAGACCGCGCGCTGCGAGCGCGCGCGCGCGCGCTTCTGCTGGA 600

Qy 658 CGGATTTCTGGGGAAGCGGAGCTCCGAGGGGTGCGAGCGCGCGCGCGCGCGTCCGCGT 717

Db 601 CGGATTTCTGGGGAAGCGGAGCTCCGAGGGGTGCGAGCGCGCGCGCGCGCGTCCGCGT 660

Qy 718 GCCCGCGCGCTTGGGCTTGGCTGAGTGGCGCGCGCGCGCGCGCGCGCGCTGCTG 777

Db 661 GCCCGCGCGCTTGGGCTTGGCTGAGTGGCGCGCGCGCGCGCGCGCGCGCTGCTG 720

Qy 778 CGTGTAAACGCTTAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTGCGACCC 837

Db 721 CGTGTAAACGCTTAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTGCGACCC 780

RESULT 8

AAZ98064

ID AAZ98064 standard; cDNA; 1050 BP.

XX AC AAZ98064;

XX AC

DT 09-MAY-2000 (first entry)

XX AC

DE Human secreted protein encoding nucleotide sequence SEQ ID NO:58.

XX AC

KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;

KW antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;

KW tumour; neurodegenerative disorder; developmental abnormality; allergy;

KW foetal deficiency; blood disorder; immune system disorder; arthritis;

KW autoimmune disease; hepatic disease; renal disease; inflammation;

KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;

KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;

KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;

KW reproductive disorder; gastrointestinal disorder; respiratory disorder;

KW metabolic disorder; food additive; preservative; ss.

XX OS Homo sapiens.

XX OS

PN W0200004140-A1.

XX PN

PD 27-JAN-2000.

XX PD

XX 14-JUL-1999; 99WO-US15849.

PF 14-JUL-1999; 99WO-US15849.

XX PF

PR 15-JUL-1998; 98US-0092921.

XX PR

PR 15-JUL-1998; 98US-0092922.

PR 15-JUL-1998; 98US-0092956.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA

PI Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;

PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;

PI Mucenski M, Endress CA, Soppet DR;

XX PI

DR WPI: 2000-161128/14.

DR P-PSDB; AAY87111.

XX DR

XX New isolated human genes, useful for diagnosis and treatment of, e.g.

PT cancers, neurological or blood disorders

PT

PS Claim 1; Page 339; 494pp; English.

XX PS

CC The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the

CC human secreted proteins given in AAY87064 to AAY87223. Human secreted

CC protein can have activities based on the tissues and cells the genes are

CC expressed in. Examples of activities include: cytostatic;

CC immunosuppressive; antiinflammatory; nontropic; neuroprotective; and

CC antiallergic. The polynucleotides and their corresponding secreted

CC polypeptides are useful for preventing, treating or ameliorating medical

CC conditions, e.g. by protein or gene therapy. Also pathological conditions

CC can be diagnosed by determining the amount of the new polypeptides in a

CC sample or by determining the presence of mutations in the new

CC polynucleotides. Human secreted protein s and their polynucleotides can

CC be used for developing products for the diagnosis or treatment of cancer,

CC tumours, neurodegenerative disorders, developmental abnormalities and

CC foetal deficiencies, blood disorders, diseases of the immune system,

CC autoimmune diseases, hepatic and renal disease, inflammation,

CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,

CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,

CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,

CC cardiovascular disorders, reproductive disorders, gastrointestinal

CC disorders, respiratory disorders and metabolic disorders. The

CC proteins or polynucleotides can also be used as food additives or

CC preservatives. The proteins are also useful for identifying their

CC binding partners. AAZ98008 to AAZ98016 and AAY87063 are sequence used in

CC the exemplification of the present invention.

XX CC

SQ Sequence 1050 BP; 172 A; 398 C; 344 G; 136 T; 0 other;

Query Match 59.6%; Score 583; DB 21; Length 1050;

Best Local Similarity 99.8%; Pred. No. 3.1e-250;

Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 346 GAGGCTGAGGATCAGAGGCGCGGCTCTGCGCAGGTGCTGCGGCGCGCCCC 405

Db 339 GAGGCTGAGGATCAGAGGCGCGGCTCTGCGCAGGTGCTGCGGCGCGCCCC 398

Qy 406 CGCAACTCTATCGGCTCTGGCTTGGAGCAGACACCGCGCGCTGGAGCGACGTC 465

Db 399 CGCAACTCTATCGGCTCTGGGCTTGGAGCAGACACCGCGCGCTGGAGCGACGTC 458

Qy 466 GCTCGGCTCTGCTCCCGCGCGCGCTTACCTCGCGCGCTAGCAGCGCGCTTGTCCCC 525

Db 459 GCTCGGCTCTGCTCCCGCGCGCGCTTACCTCGCGCGCTAGCAGCGCGCTTGTCCCC 518

Qy 526 GCGGCGCTCCCGCGCGCGGCTTCCGACCGCGCGCGCGCGCGCTTACGACGCGCGCGG 585

Db 519 GCGGCGCTCCCGCGCGCGGCTTCCGACCGCGCGCGCGCGCGCTTACGACGCGCGCGG 578

Qy 586 GGCGCGGATCTGAGGAGGCGGCGGACGACACCGCGCGCTGGACCGCGAGCTTTCAGG 645

Db 579 GGCGCGGATCTGAGGAGGCGGCGGACGACACCGCGCGCTGGACCGCGAGCTTTCAGG 638

Qy 646 TACTTGTGGGAGGATTCCTTGGGGAAGCGCGGACTTCCGAGGGGGTGGCAGCGCGCGCG 705

Db 639 TACTTGTGGGAGGATTCCTTGGGGAAGCGCGGACTTCCGAGGGGGTGGCAGCGCGCGCG 698

Qy 706 CGGCTCGCGCGCTGCGCGCGGACGAGATGTGGGCTCTGAGCTGCCCTGAGGGGCTGCTG 765

```
Db 699 CGCCTCGCGCTGCGCGGACACAGATGGGGCTCTGAGTGCCTCCCTGAGGGCGTGTG 758
QY 766 GGGGCGCTGCTGGTGTGAACGCTTAGAGACCCCGCGCGCCAGGTGCTGCACGCCGC 825
Db 759 GGGGCGCTGCTGGTGTGAACGCTTAGAGACCCCGCGCGCCAGGTGCTGCACGCCGC 818
QY 826 CTCTTGGCCACCTGAGCACTGCGCGGATCCGTGACCTTGGAGCCAGAGTGCCTCCG 885
Db 819 CTCTTGGCCACCTGAGCACTGCGCGGATCCGTGACCTTGGAGCCAGAGTGCCTCCG 878
QY 886 CCATCCCGCCACAGCACTGCTCCCGCGCAGCAGCTCCAGAGCACTTACCCCGGCCAGC 945
Db 879 CCATCCCGCCACAGCACTGCTCCCGCGCAGCAGCTCCAGAGCACTTACCCCGGCCAGC 938
QY 946 CAGCCCTCTACCCGAGGATCCCTACCCCTGCG 979
Db 939 CAGCCCTCTACCCGAGGATCCCTACCCCTGCG 972

RESULT 9
AADI1677
ID AADI1677 standard; cDNA; 1050 BP.
XX
AC AADI1677;
XX
DT 24-SEP-2001 (first entry)
XX
DE Human secreted protein-encoding gene 48 cDNA clone HKGDL36, SEQ ID NO:58.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnery; binding partner identification;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT CDS 55..501
FT /tag= a
FT /product= "Human secreted protein precursor"
FT sig_peptide 55..153
FT /tag= b
FT mat_peptide 154..498
FT /tag= c
FT /product= "Mature human secreted protein"
XX
WO200151504-A1.
XX
19-JUL-2001.
XX
12-JAN-2001; 2001WO-US00911.
XX
13-JAN-2000; 2000US-0482273.
XX
(HUMA-) HUMAN GENOME SCI INC.
XX
Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Muscenski M, Ebner R;
XX
WPI; 2001-425865/45.
DR P-PSDB; AAE06088.
XX
Isolated nucleic acid molecule encoding a human secreted protein is
used in preventing, treating or ameliorating a medical condition -
XX
```

```
PS
XX
Claim 1; Page 706; 864pp; English.
AADI1630-AADI1721 represent cDNAs corresponding to 71 human secreted
protein genes, and AAE06041-AAE06132 represent the proteins they encode.
AAE06133-AAE06205 represent human secreted protein fragments.
The secreted proteins and their genes are useful for preventing, treating
or ameliorating medical conditions, e.g., by protein or gene therapy.
Pathological conditions can be diagnosed by determining the amount of the
new protein in a sample or by determining the presence of mutations in
the new genes. Specific uses are described for each of the 71 genes,
based on the tissues in which they are most highly expressed, and include
developing products for the diagnosis or treatment of proliferative
disorders, cancer, tumours, foetal and developmental abnormalities,
haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
diseases (e.g., rheumatoid arthritis), inflammation, allergies,
neurological disorders (e.g., Alzheimer's disease, parkinson's disease),
cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
angogenic disorders, kidney disorders, gastrointestinal disorders,
pregnancy-related disorders, endocrine disorders, and infectious. The
proteins can also be used to aid wound healing and epithelial cell
proliferation, to prevent skin aging due to sunburn, to maintain organs
before transplantation, for supporting cell culture of primary tissues,
to regenerate tissues, to identify their cognate ligands or binding
partners, and in chemotaxis, and can be used as a food additive or
preservative to modify storage properties. Antibodies specific for a
protein of the invention can be used in alleviating symptoms associated
with the disorders mentioned above, and in diagnostic immunoassays e.g.,
radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
The present sequence represents a human secreted protein-encoding cDNA of
the invention.
SQ Sequence 1050 BP; 172 A; 398 C; 344 G; 136 T; 0 Other;
Query Match 59.6%; Score 583; DB 22; Length 1050;
Best Local Similarity 99.8%; Pred. No. 3.1e-250;
Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 346 GAGGCTGAGGATCAGCAGCGCGCTCTGGCGCAGCTGCTGCGCGCGCCCC 405
Db 339 GAGGCTGAGGATCAGCAGCGCGCTCTGGCGCAGCTGCTGCGCGCGCCCC 398
QY 406 CGCAACTCTGATCCGGCTTGGGCTTGGAGCAGACCCCGAGCGCGCTGAGCGCAGCTC 465
Db 399 CGCAACTCTGATCCGGCTTGGGCTTGGAGCAGACCCCGAGCGCGCTGAGCGCAGCTC 458
QY 466 GCTCGCGCTGCTGCTCGCGCGCGCTTGGAGCAGCTGCTGCGCGCGCCCC 525
Db 459 GCTCGCGCTGCTGCTCGCGCGCGCTTGGAGCAGCTGCTGCGCGCGCCCC 518
QY 526 GCGCCCGTCCCGCGCGCGCTTGGAGCAGCTGCTGCGCGCGCCCCCGCG 585
Db 519 GCGCCCGTCCCGCGCGCGCTTGGAGCAGCTGCTGCGCGCGCCCCCGCG 578
QY 586 GGGCCGGATGCTGAGGAGCGAGCAGACACCCGAGCTGAGACCCCGAGCTGTTGAGG 645
Db 579 GGGCCGGATGCTGAGGAGCGAGCAGACACCCGAGCTGAGACCCCGAGCTGTTGAGG 638
QY 646 TACTTCTGGGACGATTTCTTGGGGAAGCGCGAGCTCCGAGGGGTGGCAGCCCGCGC 705
Db 639 TACTTCTGGGACGATTTCTTGGGGAAGCGCGAGCTCCGAGGGGTGGCAGCCCGCGC 698
QY 706 CGCCTCGCGCGTCCCGCGCGAGCAGCTGTTGGGCTCTGAGCTGCCCGCTGAGGCGCTGCTG 765
Db 699 CGCCTCGCGCGTCCCGCGCGAGCAGCTGTTGGGCTCTGAGCTGCCCGCTGAGGCGCTGCTG 758
QY 766 GGGGCGCTGCTGCTGTGAACGCTTAGAGACCCCGCGCGCCAGGTGCTGCACGCCGC 825
Db 759 GGGGCGCTGCTGCTGTGAACGCTTAGAGACCCCGCGCGCCAGGTGCTGCACGCCGC 818
QY 826 CTCTTCCCACTTACCACTGCGCGGATCCCGTGCACCTTGGAGCCAGAGTGCCTCCCG 885
Db 819 CTCTTCCCACTTACCACTGCGCGGATCCCGTGCACCTTGGAGCCAGAGTGCCTCCCG 878
```

QY 886 CCATCCGCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGCCAGC 945
|||||
Db 879 CCATCCGCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGCCAGC 938
QY 946 CAGCCCTCTACCCGAGGATCCCTACCCCTGGC 979
|||||
Db 939 CAGCCCTCTACCCGAGGATCCCTACCCCTGGC 972

RESULT 10
ABK69773
ID ABK69773 standard; cDNA; 1050 BP.
XX
AC ABK69773;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein gene 48.
XX
KW Human; ss; gene; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; ophthalmological; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US29871.
XX
PR 25-SEP-2000; 2000US-234925P.
PR 12-JAN-2001; 2001WO-US00911.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress CA, Mucenski M, Ebner R;
XX
DR WPI: 2002-362489/39.
DR P-PSDB: ABG33910, ABG34000.
XX
XX
PT Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT the polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT asthma
XX
PS Claim 1; Page 1189; 1478pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (or its
XX fragment, homologue complement or allelic variant) encoding a human
XX secreted protein (and its fragment, domain, epitope, variant, secreted
XX form and species variant). Also included are a recombinant vector
XX comprising the nucleic acid, a recombinant host cell comprising the
XX vector, an antibody against the secreted protein, a recombinant host cell
XX that expresses the secreted protein and a method of identifying a binding
XX partner of the secreted protein. The nucleic acid and protein are used to
XX prevent, diagnose, treat or ameliorate a medical condition in e.g.
XX humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
XX for example autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi and
XX ocular disorders e.g. corneal infection. Many other diseases and

CC disorders are listed in the specification. The polypeptides can also be
CC used to aid wound healing an epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence encodes a novel human secreted protein of the invention.
XX
SQ Sequence 1050 BP; 172 A; 398 C; 344 G; 136 T; 0 other;

Query Match 59.68; Score 583; DB 24; Length 1050;
Best Local Similarity 99.8%; Pred. No. 3.1e-250;
Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 346 GAGGCTGAGGATCAGCAGCGCGCTCTGGCGCAGCTGCTGGCGTCTGGGGGCGCCGCC 405
|||||
Db 339 GAGGCTGAGGATCAGCAGCGCGCTCTGGCGCAGCTGCTGGCGTCTGGGGGCGCCGCC 398
|||||
QY 406 CGCAACTCTGATCGGCTCTGGGCTTGGACGACGACCCCGCCTGACGCGCAGCTC 465
|||||
Db 399 CGCAACTCTGATCGGCTCTGGGCTTGGACGACGACCCCGCCTGACGCGCAGCTC 458
|||||
QY 466 GCTCGGCTCTGCTCCGCGCGCGCTTGGACCTGCGCGCCTAGCAGCCAGCTGTGCC 525
|||||
Db 459 GCTCGGCTCTGCTCCGCGCGCGCTTGGACCTGCGCGCCTAGCAGCCAGCTGTGCC 518
|||||
QY 526 GCGCCGCTCCCGCGCGCGCTCCGACCCCGCGCGCTACGACGACGCGCCCGCG 585
|||||
Db 519 GCGCCGCTCCCGCGCGCGCTCCGACCCCGCGCGCTACGACGACGCGCCCGCG 578
|||||
QY 586 GGCCCGGATGCTAGGAGGCGAGGACGACACCCCGACGCTGGACCCCGAGCTGTGAGG 645
|||||
Db 579 GGCCCGGATGCTAGGAGGCGAGGACGACACCCCGACGCTGGACCCCGAGCTGTGAGG 638
|||||
QY 646 TACTTGTGGGACGGATCTTTCGGGAGCGCGGACCTCCGAGGGGTGGCAGCCCGCGC 705
|||||
Db 639 TACTTGTGGGACGGATCTTTCGGGAGCGCGGACCTCCGAGGGGTGGCAGCCCGCGC 698
|||||
QY 706 CGCCTCGCGCTGCGCGCGCGACGACGATGTGGCTCTGAGCTGCCCCCTGAGGGCGTGTG 765
|||||
Db 699 CGCCTCGCGCTGCGCGCGCGACGACGATGTGGCTCTGAGCTGCCCCCTGAGGGCGTGTG 758
|||||
QY 766 GGGCGCTGTGCTGTGAACGCTTAGAGACCCCGCGCGCGCGCTGCGACGCGCGC 825
|||||
Db 759 GGGCGCTGTGCTGTGAACGCTTAGAGACCCCGCGCGCGCGCTGCGACGCGCGC 818
|||||
QY 826 CTCTTGCACCTCAGCAGCTGCGCGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCG 885
|||||
Db 819 CTCTTGCACCTCAGCAGCTGCGCGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCG 878
|||||
QY 886 CCATCCGCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGCCAGC 945
|||||
Db 879 CCATCCGCGCCACAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGCCAGC 938
|||||
QY 946 CAGCCCTCTACCCGAGGATCCCTACCCCTGGC 979
|||||
Db 939 CAGCCCTCTACCCGAGGATCCCTACCCCTGGC 972
|||||

RESULT 11
AAI58162
ID AAI58162 standard; cDNA; 919 BP.
XX
AC AAI58162;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 365.
XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

PS Example 2; Page 1404-1406; 1478pp; English.

XX The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC for example autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. Many other diseases and
CC disorders are listed in the specification. The polypeptides can also be
CC used to aid wound healing an epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence represents a ds DNA fragment of the gene for a novel human
CC secreted protein of the invention.

XX Sequence 4503 BP; 783 A; 1181 C; 1314 G; 1225 T; 0 other;

Query Match 38.3%; Score 375; DB 24; Length 4503;

Best Local Similarity 99.6%; Pred. No. 1.5e-137; Indels 0; Gaps 0;
Matches 475; Conservative 0; Mismatches 2;

Qy 170 AGGAGCCCGCGGCTAGACGAGCGTCTCGCCCTTGGCTGAGACTGGCGCTCTCGCC 229
Db 3254 AGGAGCCCGCGGCTAGACGAGCGTCTCGCCCTTGGCTGAGACTGGCGCTCTCGCC 3313
Qy 230 GCTTCGCGGCTAGTGCCTCCGAGGTAGGCGGCGGCGGCGGTGAGAGTGGCGCGG 289
Db 3314 GCTTCGCGGCTAGTGCCTCCGAGGTAGGCGGCGGCGGCGGTGAGAGTGGCGCGG 3373
Qy 290 CGCTGGCGCATCTGCTGAGCGCGCACTGAGGAGCGGCGGCGGCGGCGGCGGCGG 349
Db 3374 CGCTGGCGCATCTGCTGAGCGCGCACTGAGGAGCGGCGGCGGCGGCGGCGGCGG 3433
Qy 350 CTGAGGATCAGCAGCGCGCTCTCTGGCGCAGCTGCTGCGGCTCTGGGCGCGCCCGCA 409
Db 3434 CTGAGGATCAGCAGCGCGCTCTCTGGCGCAGCTGCTGCGGCTCTGGGCGCGCCCGCA 3493
Qy 410 ACTGTATCCGGCTCTGGGCTTGGACGACGACCCCGAGCGGCGCTGACGCGAGTCTGCTC 469
Db 3494 ACTGTATCCGGCTCTGGGCTTGGACGACGACCCCGAGCGGCGCTGACGCGAGTCTGCTC 3553
Qy 470 GCGTCTGCTCCGCGCGCTTACCTCTGCGGCTGAGCAGCCAGCTTGTCCCGCGC 529
Db 3554 GCGTCTGCTCCGCGCGCTTACCTCTGCGGCTGAGCAGCCAGCTTGTCCCGCGC 3613
Qy 530 CCGTCCCGCGCGCGCTTCCGACCCCGCGGCTTACGACGAGCGGCGCGCGCGGCGC 589
Db 3614 CCGTCCCGCGCGCGCTTCCGACCCCGCGGCTTACGACGAGCGGCGCGCGCGGCGC 3673
Qy 590 CGGATGTGAGGAGCGGCGGACGAGACACCCGACGTGGACCCCGAGGTGTGTGAGGT 646
Db 3674 CGGATGTGAGGAGCGGCGGACGAGACACCCGACGTGGACCCCGAGGTGTGTGAGGT 3730

RESULT 13

ABN95203/c

ID ABN95203 standard; DNA; 301 BP.

XX

XX ABN95203;

XX

DT 13-AUG-2002 (first entry)

XX Gene #1701 used to diagnose liver cancer.
DE Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX Homo sapiens.
XX WO200229103-A2.
XX 11-APR-2002.
XX 02-OCT-2001; 2001WO-US30589.
XX 02-OCT-2000; 2000US-237054P.
XX (GENE-) GENE LOGIC INC.
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample
XX Claim 1; SEQ ID NO 1701; 298pp; English.
XX The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cystostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 301 BP; 41 A; 79 C; 131 G; 48 T; 2 other;

Query Match 23.1%; Score 226; DB 24; Length 301;

Best Local Similarity 99.6%; Pred. No. 5e-91;
Matches 276; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 703 CGCGCGCTCCGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 762
Db 301 CGCGCGCTCCGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242
Qy 763 CTGGGGGCGCTGCTGCGCTGTGAAACGCGCTAGACACCGCGCGCGCGCGCGCGCGCG 822
Db 241 CTGGGGGCGCTGCTGCGCTGTGAAACGCGCTAGACACCGCGCGCGCGCGCGCGCG 182
Qy 823 CGCTCTTGGCCACCTTGGACACTGCGCGGATCCCGTGCACCTGGGACCCAGAGTGGCC 882
Db 181 CGCTCTTGGCCACCTTGGACACTGCGCGGATCCCGTGCACCTGGGACCCAGAGTGGCC 122
Qy 883 CGCGCATCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 942
Db 121 CGCGCATCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62
Qy 943 AGCCAGCGCTCTACCCCGGAGGATCCCTACCCCGCTGGC 979
Db 61 AGCCAGCGCTCTACCCCGGAGGATCCCTACCCCGCTGGC 25


```

PR 02-MAY-2001; 200105-287724P.
XX
FA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257393/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
XX
XX Example 1; SEQ ID 14889; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridising selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at wipo.int/pub/published\_pct\_sequences.
XX
SQ Sequence 60 BP; 12 A; 16 C; 21 G; 11 T; 0 other;

Query Match 6.1%; Score 60; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 5.8e-17;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 606 AGCGAGCAGACACCCGAGCTGGACCCGAGCTGTTGAGGTACTTGTCTGGGACGGATTCT 665
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Db 1 AGCGAGCAGACACCCGAGCTGGACCCGAGCTGTTGAGGTACTTGTCTGGGACGGATTCT 60
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Search completed: April 12, 2003, 19:33:48
Job time : 286 secs

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:14:10 ; Search time 87 Seconds
(without alignments)
615.773 Million cell updates/sec

Title: US-09-803-589-6
Perfect score: 1319
Sequence: 1 MAGSPLLWGPAGGVGLLV.....RVKRLTAPQVARRLLP 260

Scoring table: BLOSUMP62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1319	100.0	260	4	Q9UHG2	Q9uhg2 homo sapien
2	1102	83.5	260	11	Q9QXU9	Q9qxu9 rattus norv
3	1082	82.0	258	11	Q91W26	Q91w26 mus musculu
4	1076	81.6	258	11	Q9QXV0	Q9qxv0 mus musculu
5	490.5	37.2	187	11	Q9ESU4	Q9esu4 mus musculu
6	146	11.1	913	16	Q8UJ51	Q8uj51 agrobacteri
7	136.5	10.3	388	4	Q9NY42	Q9ny42 homo sapien
8	136.5	10.3	441	4	Q75685	Q75685 homo sapien
9	134	10.2	553	16	Q9RSN4	Q9rsn4 deinococcus
10	132.5	10.0	612	2	Q9Z614	Q9z614 streptomyce
11	131	9.9	492	10	Q9RV89	Q9rv89 oryza sativ
12	131	9.9	1290	5	Q9NEA7	Q9nea7 leishmania
13	130.5	9.9	589	16	Q9A718	Q9a718 caulobacter
14	130	9.9	350	2	Q9EU24	Q9eu24 corynebacte
15	129.5	9.8	544	11	Q9VCG2	Q9vcg2 mus musculu
16	129	9.8	660	16	Q9RDL3	Q9rdl3 streptomyce

17	127.5	9.7	460	5	Q9GZF7	Q9gzf7 caenorhabdi
18	125.5	9.5	431	16	Q9L0J8	Q9l0j8 streptomyce
19	125.5	9.5	816	4	Q9G51	Q9g51 homo sapien
20	125.5	9.5	1340	16	Q91LH8	Q91lh8 streptomyce
21	124.5	9.4	4809	2	Q93HH0	Q93hh0 streptomyce
22	123.5	9.4	700	2	Q34003	Q34003 rhodobacter
23	123	9.3	801	5	Q23635	Q23635 caenorhabdi
24	121.5	9.2	590	16	Q9S2Q5	Q9s2q5 streptomyce
25	121	9.2	772	16	Q9AD96	Q9ad96 streptomyce
26	121	9.2	791	11	Q9Z1P7	Q9z1p7 mus musculu
27	120.5	9.1	817	4	Q96SB3	Q96sb3 homo sapien
28	119	9.0	421	5	Q9VGC1	Q9vgc1 drosophila
29	119	9.0	469	12	Q69270	Q69270 equine herp
30	119	9.0	934	16	Q9K452	Q9k452 streptomyce
31	118.5	9.0	667	16	Q9RJY5	Q9rjy5 streptomyce
32	118.5	9.0	1641	5	Q9GRZ3	Q9grz3 caenorhabdi
33	118	8.9	238	11	Q9JL0	Q9jl0 mus musculu
34	118	8.9	1240	12	Q9DWH8	Q9dwh8 rat cytomeg
35	118	8.9	1430	11	Q9VHK2	Q9vnh2 rattus norv
36	117.5	8.9	327	11	Q99JK6	Q99jk6 mus musculu
37	117.5	8.9	615	4	Q9UDW8	Q9udw8 homo sapien
38	117.5	8.9	941	16	Q9Z1W5	Q9z1w5 streptomyce
39	117	8.9	406	2	Q8VPM8	Q8vpm8 micrococcus
40	117	8.9	2157	11	Q9Z1R1	Q9z1r1 mus musculu
41	116.5	8.8	1198	11	Q9JKA7	Q9jka7 rattus norv
42	116.5	8.8	1677	11	Q70373	Q70373 mus musculu
43	116.5	8.8	2307	2	Q9AG79	Q9ag79 streptomyce
44	116.5	8.8	6146	2	Q93HJ5	Q93hj5 streptomyce
45	116	8.8	710	4	Q9H0J3	Q9h0j3 homo sapien

ALIGNMENTS

RESULT 1	
Q9UHG2	PRELIMINARY; PRT; 260 AA.
AC Q9UHG2;	
DT 01-MAY-2000 (TREMBLrel. 13, Created)	
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE PROSNAAS precursor (Granin-like neuroendocrine peptide precursor).	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RA Fricker L., McKinzie A.A., Sun J., Curran E., Qian Y., Yan L.,	
RA Patterson S.D., Courchesne P.L., Richards B., Levin N., Mzhavia N.,	
RA Devi L.A., Douglass J.;	
RT "Identification and characterization of prosnaas: a granin-like	
RT neuroendocrine peptide precursor that inhibits prohormone	
RT processing";	
RL J. Neurosci. 20:0-0(2000).	
RN [2]	
RP SEQUENCE FROM N.A.	
RC TISSUE=UTERUS;	
RA Strausberg R.;	
RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.	
DR EMBL; AF181562; AAF22643.1; -;	
DR EMBL; BC002851; AAH02851.1; -;	
KW Signal.	
FT SIGNAL 1 33 POTENTIAL.	
SQ SEQUENCE 260 AA; 27372 MW; FF8E2727284B7A5C CRC64;	
Query Match 100.0%; Score 1319; DB 4; Length 260;	
Best Local Similarity 100.0%; Pred. NO. 3.8e-74;	
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 MAGSPLLWGPAGGVGLLVLLLLGLFRPPALCARVPKPRGLSAASPPLAETGAPRRFR 60	
DB 1 MAGSPLLWGPAGGVGLLVLLLLGLFRPPALCARVPKPRGLSAASPPLAETGAPRRFR 60	

QY 61 RSVPRGEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLLRVWGAPRNSD 120
 DB 61 RSVPRGEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLLRVWGAPRNSD 120
 QY 121 PALGLDDDDPAPAAQALRALRLRDLPAALAAQLVPAPVPAALRPPVYDDGPGAPDA 180
 DB 121 PALGLDDDDPAPAAQALRALRLRDLPAALAAQLVPAPVPAALRPPVYDDGPGAPDA 180
 QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSGLPEGVIGALL 240
 DB 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSGLPEGVIGALL 240
 QY 241 RVKRLTAPQVPAARRLLPP 260
 DB 241 RVKRLTAPQVPAARRLLPP 260

RESULT 2

QYXU9 ID QYXU9 PRELIMINARY; PRT; 260 AA.
 AC QYXU9;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE ProSAAS precursor.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20098938; PubMed=10632593;
 RA Fricker L., McKinzie A.A., Sun J., Curran E., Qian Y., Yan L.,
 RA Patterson S.D., Courchesne P.L., Richards B., Levin N., Mzhavia N.,
 RA Devi L.A., Douglass J.
 RT "Identification and characterization of proSAAS, a granin-like
 RT neuroendocrine peptide precursor that inhibits prohormone
 RT processing."
 RL J. Neurosci. 20:639-648(2000).
 DR EMBL; AF181561; AAF22642.1; -.
 KW Signal.
 FT SIGNAL
 SQ SEQUENCE 1 33 POTENTIAL.
 SEQUENCE 260 AA; 27414 MW; C84F688BDEB5313E CRC64;

Query Match 83.5%; Score 1102; DB 11; Length 260;
 Best Local Similarity 84.6%; Pred. No. 8.6e-61;
 Matches 220; Conservative 8; Mismatches 32; Indels 0; Gaps 0;
 QY 1 MAGSPLMGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPLAETGAPRRFR 60
 DB 1 MAGSPLMGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPLAETGAPRRFR 60
 QY 61 RSVPRGEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLLRVWGAPRNSD 120
 DB 61 RSVPRGEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLLRVWGAPRNSD 120
 QY 121 PALGLDDDDPAPAAQALRALRLRDLPAALAAQLVPAPVPAALRPPVYDDGPGAPDA 180
 DB 121 PALGLDDDDPAPAAQALRALRLRDLPAALAAQLVPAPVPAALRPPVYDDGPGAPDA 180
 QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSGLPEGVIGALL 240
 DB 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSGLPEGVIGALL 240
 QY 241 RVKRLTAPQVPAARRLLPP 260
 DB 241 RVKRLTAPQVPAARRLLPP 260

RESULT 3

QY1W26 ID QY1W26 PRELIMINARY; PRT; 258 AA.

QY1W26;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Similar to granin-like neuroendocrine peptide.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COLON;
 RA Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC012263; AAH12263.1; -.
 SQ SEQUENCE 258 AA; 27270 MW; 4197C8B077A20A22 CRC64;

Query Match 82.0%; Score 1082; DB 11; Length 258;
 Best Local Similarity 83.5%; Pred. No. 1.5e-59;
 Matches 217; Conservative 10; Mismatches 31; Indels 2; Gaps 1;
 QY 1 MAGSPLMGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPLAETGAPRRFR 60
 DB 1 MAGSPLMGPRAGGVGLLVLLGLFRPPPPALCARPVKEPRGLSAASPLAETGAPRRFR 60
 QY 61 RSVPRGEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLLRVWGAPRNSD 120
 DB 61 RSVPRGEAAGAVQELARALAHLEAERQERARAQAEDQOARVLAQLLRVWGAPRNSD 120
 QY 121 PALGLDDDDPAPAAQALRALRLRDLPAALAAQLVPAPVPAALRPPVYDDGPGAPDA 180
 DB 121 PALGLDDDDPAPAAQALRALRLRDLPAALAAQLVPAPVPAALRPPVYDDGPGAPDA 180
 QY 181 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSGLPEGVIGALL 240
 DB 179 EAGDETDPDPELLRYLLGRILAGSADSEGVAAAPRRLRAADHDVGSGLPEGVIGALL 240
 QY 241 RVKRLTAPQVPAARRLLPP 260
 DB 239 RVKRLTAPQVPAARRLLPP 258

RESULT 4

QYXV0 ID QYXV0 PRELIMINARY; PRT; 258 AA.
 AC QYXV0;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE ProSAAS precursor.
 GN PCSK1N OR SAAS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20098938; PubMed=10632593;
 RA Fricker L., McKinzie A.A., Sun J., Curran E., Qian Y., Yan L.,
 RA Patterson S.D., Courchesne P.L., Richards B., Levin N., Mzhavia N.,
 RA Devi L.A., Douglass J.
 RT "Identification and characterization of proSAAS, a granin-like
 RT neuroendocrine peptide precursor that inhibits prohormone
 RT processing."
 RL J. Neurosci. 20:639-648(2000).
 DR EMBL; AF181560; AAF22641.1; -.
 DR MGI; MGI:1353431; Pcskn.
 DR InterPro; IPR002965; P_rich_extensn.
 DR PRINTS; PRO1217; PRICEXTENS.
 KW Signal.
 FT SIGNAL
 SQ SEQUENCE 1 33 POTENTIAL.
 SEQUENCE 258 AA; 27285 MW; 07452A460E868CF6 CRC64;

RESULT 7

Query Match 10.2%; Score 134; DB 16; Length 553;
Best Local Similarity 25.0%; Pred. No. 0.72;
Matches 70; Conservative 24; Mismatches 108; Indels 78; Gaps 11;

QY 24 GLFRPPALCARPVKEPRG-----LSAASP-----LAET 53
DB 32 GTLAPEAPVQIRPVRRPAGQVPTTSIALSGVEPPAPAEVQVVRAYRDPDSRTPTAAEEP 91
QY 54 GAPRRFRSVPGRGAGAVQELARALAHLEAEERQER-----AEQAEDQOAR 104
DB 92 VAPTRTSKARARPTPAETVPSAPVQAPVAEAEPEKPRPTKAAAPVPAASPAETEAA 151
QY 105 VLAQLLRWAGAPRNSDPAIGDD---DPDAPAAQALARALRLDPAALAAQ-----LVPAP 158
DB 152 PAS-----APRSGWLSALDLWKEPEAPAPRREAPARQVQERSTATQTQVIOAAP 204
QY 159 VPAALAPRPVYDDGPAGDAEB---AGDETDPDPELLRYLLGRILAGSADSEGVAAP 215
DB 205 KPAPA--PQP---DSDSLPRTIQEALASDLPLDPLVELLERLWEQEAQAEQEEPAP 259
QY 216 RRLRRAADHDVGSPELPPGVGLGALLRYKRLTPAQPVPAR 255
DB 260 ----RAPVRPVAALSP-----APSPAAPP 281

RESULT 10
Q92614 ID Q92614 PRELIMINARY; PRT; 612 AA.
AC Q92614;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Dihydrolipoamide acetyltransferase.
GN PDHB.
OS Streptomyces seoulensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=73044;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99078078; PubMed=9858775;
RA Youn H., Kwak J., Youn H.D., Hah Y.C., Kang S.O.;
RT "Lipoamide dehydrogenase from streptomyces seoulensis: biochemical and genetic properties";
RL Biochim. Biophys. Acta 1388:405-418(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Youn H., Kang S.-O.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 LIPOYL-BINDING DOMAINS.
DR HSP; P11961; 2PDD.
DR InterPro; IPR001078; 2oxoacid dh.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR004167; E3-binding.
DR InterPro; IPR003016; Lipoyl.
DR InterPro; IPR002965; P-rich.extensn.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00198; 2-oxoacid dh; 1.
DR Pfam; PF00364; biotin_lipoyl; 2.
DR Pfam; PF02817; e3-binding; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRINTS; PR01574; TUBBYPROTEIN.
DR ProDom; PD001115; 2oxoacid dh; 1.
DR PROSITE; PS00189; LIPOYL; 2.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Lipoyl; Transferase.
SQ SEQUENCE 612 AA; 61368 MW; B804109089943056 CRC64;

Query Match 10.0%; Score 132.5; DB 2; Length 612;
Best Local Similarity 27.0%; Pred. No. 0.99;
Matches 70; Conservative 21; Mismatches 117; Indels 51; Gaps 10;

QY 29 PPALCARPVKEPRGLSASPLAET---GAPRRFRSVPGRGAGAVQELARALAHLEA 85
DB 121 PTAESAAGSGSAECTDVLVLPALGSEVTEGTVRLKSV--GDSVEADEPILLEVSTDKVDT 178
QY 86 ERQERARAEQ---AEDQOARVLAOLLRVWGAAPRNSDPAIGLDDDDPDAPAAQLARALLR 142
DB 179 EIPAPTSGTLLLEIVGDETAEGAKLAVIGAA--GAAPA-AAPEAPKAPAAEAPAAPAP 235
QY 143 ARLDPAALAAQLVPAPVPAALRPRP-----PVYDDGPA-----GP 178
DB 236 AAP 295
QY 179 DAEPAGDETPDPELLRYLLGRILAGSADSEGVAAPRRRLRRAADHDVGSPELPPGVGLA 238
DB 296 TATQATDEGAVTBP-LVRKLAASESSVDLASVKGTVGGRIRK-----QDVIAA 342
QY 239 LLRVKRLTPAP---QVPA 254
DB 343 AEAATAAAAPAPAAAAQAPA 361

RESULT 11
Q8RV89 ID Q8RV89 PRELIMINARY; PRT; 492 AA.
AC Q8RV89;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 52.5 kDa protein.
GN OSUNBA0095C06.3 OR OSUNBA0091J06.20.
OS Oryza sativa (Rice), and
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriarthroidae; Oryzae; Oryza.
OX NCBI_TaxID=4530; 39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Du H., Minx P., de la Bastide M., Nascimento L., Spiegel L.,
RA Preston R., Kirchoff K., Kuit K., Baker J., Vil M.D., Zutavern T.,
RA Santos L., Cunniss D.M., Rodriguez S., Miller B., Ballija V., Shah R.,
RA Bahret A., King L., O'Shaughnessy A., Palmer L., Dedhia N.,
RA McCombie W.R.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
OSJNBa0095C06, from Chromosome 10, complete sequence";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA McCombie W.R., de la Bastide M., Spiegel L., Preston R., Ferraro K.,
RA Kuit K., Nascimento L., Zutavern T., Ballija V., Bell M., Baker J.,
RA Santos L., Miller B., Katzenberger F., Muller S., King L., Yang C.,
RA Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
OSJNBa0091J06, from chromosome 10, complete sequence";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC079935; AAM08531.1;
DR EMBL; AC113338; AAM08664.1;
KW Hypothetical protein.
SQ SEQUENCE 492 AA; 52539 MW; 23ABC4531C2587E6 CRC64;

Query Match 9.9%; Score 131; DB 10; Length 492;
Best Local Similarity 31.4%; Pred. No. 0.98;
Matches 81; Conservative 11; Mismatches 108; Indels 58; Gaps 13;

QY 18 LVLLLLGLFRPPALCARPVKEPRGL---SAASPPLAETGAPRRFRSV-PRGEAAGAVQ 73
DB 137 LATLLPRLLCPCPRACASPPPPRLRLPATLASPP-PELLRPRARLVSAPRSRAAPAA 195
QY 74 ELARALAHLEAEERQERARAEQAEQOQARVLAQLLRWGAAPRNSDPAIGLDDDDPDAPA 133
DB 196 RLLRSTCAALPHRRRLASSTPASRLAATLATLLPRLLR---CPHRACPA-----SPPPP- 246

```

QY 134 AQLARALLRARDLPAALAAQLVPAPVPAALRRPPVYDDGPAAGDAEAGDETDFVDPE 193
DB 247 -----RLLR-----PATLSP-----PPRLRPRARLSVAPRSRAAPAA-----R 283
QY 194 LLRYLLG-----RLAGSADSEGVAA-----PRLRRAADHDVGSSELPPEGVGLALLRV 242
DB 284 LLRSTCAALPHRRRLASSPAPSLAATLATLLPRLRCRACAPASPPPPR-----LLRP 338
QY 243 KRLETPAQVPARRLLPP 260
DB 339 ATLASPPP-----RLLRP 351

RESULT 12
Q9NEA7 PRELIMINARY; PRT; 1290 AA.
AC Q9NEA7:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical protein 136.2 kDa protein.
GN L5515.05.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Masny D., Purnelle B., Goffeau A., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL161414; CAB7679.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1290 AA; 136159 MW; 6A6F8B395CE6818 CRC64;

Query Match 9.9%; Score 131; DB 5; Length 1290;
Best Local Similarity 26.6%; Pred. No. 2.6;
Matches 69; Conservative 23; Mismatches 95; Indels 72; Gaps 12;

QY 27 RPPALCARPVKEPR-GLSAASPPLAETGAP-----RRF 59
DB 95 RDPALLHTVHAQQRGVSAAGKPRARHSAPVGPDRSTDASDFSTVPFSLRLEVERRF 154
QY 60 RR-SVPGRGAAGVQ-ELARALAHLLAEQERARAEQAEQARVLAQLLRVWGAP 116
DB 155 RQMKVHGLTAASQVDIMERSVOLFOER--RTRFAQALAEQDAIWA---RWV--P 207
QY 117 RNSDPALGLDDPDAPAAQLARALLRDLPAALAAQLVPAPVPAALRRPPVYDDGPA 176
DB 208 GCSLEK---TETEAUHQAAASARVAMPASRRSDSSATPSASTAISPPPPA---SPT 261
QY 177 GPDAEAGDETPDVPDLLRYLLGRLAGSADSEGVAPRRLRRRAADHDVGSSELPPEGV 236
DB 262 G-----FLKKDIL-----DTYSKPRELREAS-----GGSRSAAAAA 292
QY 237 GALLRVKRLTPAQVPAR 255
DB 293 PAASKATKMPKGVPLAPAR 311

RESULT 13
Q9A718 PRELIMINARY; PRT; 589 AA.
AC Q9A718;

```

```

DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein CC1908.
GN CC1908.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005864; AAK23883.1; -.
DR TIGR; CC1908; -.
DR PRINTS; PR01656; VACCYTOTOXIN.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 589 AA; 59269 MW; A7E52050C891B98A CRC64;

Query Match 9.9%; Score 130.5; DB 16; Length 589;
Best Local Similarity 29.8%; Pred. No. 1.3;
Matches 82; Conservative 15; Mismatches 91; Indels 87; Gaps 15;

QY 45 AASPPLAETGAPRRFRSVPGRGEAAGAVQELARAL-AHLLEAEQERARA----- 93
DB 182 AAAPPSAVGAAP-----PWVAEGDLKALWLRLSALSLAEVSSGAVAPHVGVADG 237
QY 94 -----EAQEAEDQARVLAQLLRVWGAPRNSDPALG--LDLDDPDAPAA---QLARALLRA 143
DB 238 GADGPDPOCPMTHQARANGSSSSA-GAMASSPPAQSRLLAADAPAPATSSPIARAL--- 293
QY 144 RLDPAAALAAQLVPAPVPAALRRPPVYDDGPAAGDAEAGDETDFVDPEL-LRY----- 197
DB 294 -----PAGAPVDAEDPPPARVGAPPSG-----SPTTHADPELAVRFGAFVA 335
QY 198 -----LLGRILAGSA-----DSEG-----VAAAPRRLRR-----AARDVGS 228
DB 336 PPKTAERSPKGVAGSLPALVQLPEPRDGEPPATEPAEPRLMTRGYCAPVEDVRS 395
QY 229 ELPPPEGVGLALLRVKRL-TP-----APQVPARRLL 258
DB 396 KTPPPPYAGGPTAGORPEPTPSLSWAPEAMARRLL 430

RESULT 14
Q9EU24 PRELIMINARY; PRT; 350 AA.
AC Q9EU24;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Hypothetical 37.5 kDa protein.
GN ORF31.
OS Corynebacterium equi (Rhodococcus equi).
OC Plasmid PREAT701 (p33701), and Plasmid virulence.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=43767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC33701; PLASMID=PREAT701 (p33701);
RA Takai S., Sekizaki T., Kakuda T., Nakamura M., Suzuki K., Ogino N.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN-ATCC33701, AND 103; PLASMID-PREAT701 (P33701), AND VIRULENCE;
RX PubMed=11083803;
RA Takai S., Hines S.A., Sekizaki T., Nicholson V.M., Alperin D.A.,
RA Osaki M., Takamatsu D., Nakamura M., Suzuki K., Ogino N., Kakuda T.,
RA Dan H., Prescott J.F.;
RT "DNA sequence and comparison of virulence plasmids from *Rhodococcus*
RT equi ATCC 33701 and 103.";
RL Infect. Immun. 68:6840-6847(2000).
RL EMBL; AP001204; BAB16640.1; -;
DR EMBL; AF116907; AAG21734.1; -;
DR Hypothetical protein; Plasmid.
SQ SEQUENCE 350 AA; 37537 MW; 56A73F73318B62C3 CRC64;

Query Match 9.9%; Score 130; DB 2; Length 350;
Best Local Similarity 28.4%; Pred. NO. 0.8;
Matches 79; Conservative 20; Mismatches 89; Indels 90; Gaps 14;

Qy	27	RPPPALCARVPKEPRLGSLAASPLLATGP-	-----RRFRRSVP	64
Db	12	RPQARAIDP--GPDGAAAPQPDWAKPGRPAHEQAHQDVVDVAGFUSKVDARAQGRSVP	69	
Qy	65	R-----GEAAGA-----VOELARAL-----	AHLEAEQERARAEQAQAE	100
		:	:	
Db	70	KPQPMREALGAFHRSPTDQTRVQVLAATAARHLEGAQSANARQWERLARENDRIDALR	129	
Qy	101	-----QQAARVLAQLLRVWVGAPRNSDPALGLDDDDPA--	PAAQIARALLRAR-LD	146
Db	130	ROOHHRQEVAYQAARMSPEQIQORDAQRASASET---	ESDKRAFDAAAAIALGVYAARGLP	186
Qy	147	P-----AALAQL--VPAPYMAAL--RPRPPVYDGGPAGPDAAE--	AGDETDPDPELLR	196
Db	187	PFDRLAALGAELHDVPAPEPTTEVDDEEPTADVQSDPSAPDPQGGPAADEAVDV	-----	240
Qy	197	YLLGRILAGSADSEGVAAAPRRILRRADHDVSGELPPEG	234	
Db	241	-----SADLEPVATPEAGGASAPSRIGPELERAG	269	

RESULT 15

Q8VCG2	PRELIMINARY;	PRT;	544 AA.
ID	Q8VCG2		
AC	Q8VCG2;		
DC	01-WAR-2002 (TrEMBLrel. 20, Created)		
DT	01-WAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-WAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Hypothetical 59.8 kDa protein.		
ES	Hypothesical (Mouse).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid=10090;		
	[1]		
RP	SEQUENCE FROM N.A.		
RN			
RC	TISSUE=Salivary gland;		
RC	Strasbourg R.;		
RL	Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; BC019977; AAH19977.1; -.		
DR	Hypothetical protein.		
KW	SEQUENCE 544 AA. 59840 MW;		
SQ	SEQUENCE 544 AA. 59840 MW;		

Query Match 9.8%; Score 129.5; DB 11; Length 544;
Best Local Similarity 25.9%; Pred. No. 1.3;
Matches 74; Conservative 26; Mismatches 117; Indels 69; Gaps 13.

OY	2	AGSPILWGPACGVLIVLLGLFRPPD	PALCA	RVPKVEGRGLSASAPPIAETGA	PRFRFR	61
		:	:	:		
Db	39	AGSLAPGHPD-----	-RTHPS-AAVPC-PR--SASAPLTPC	SASSAH	81	
			: :	:		
OY	62	SVPRGE-----	AAGAVOELARALAHLLEAEQERARAEQAEDQ	QARVLAQ----	LLR	111
			: :	:		
Db	82	CTSGSERPERRKPGAGSPALARRLEATPVQKEKKDERENEKSA--	LAREN	NLK	139	
			: :	:		
OY	112	VWGAPNSDPALGLDD-----	PDAPAAQLARALLRARLDPPALAAQLVPA	PVPAAA	163	

```

Db      140 RQLPASIIRPLSTGSELSPKSKARSPSPTTWHR-----PASPCPSPGGHA 187
Qy      164 LRPRPPYDDGPAGPDA-----EEADETPDVPDELLRYLLGRILLAGSADSGVAA 214
Db      188 LPPKP---SPRGTTASPKGYRRKEEAKESPSGPCDNHKRSR----AAEEKEPAA 239
Qy      215 PRRLRAADHDVGSELPPEGVGLGALLRVKKLETTPAQPVPARRLLPP 260
Db      240 P---ASPAPSPVSPTPAOPKQKSQSQTIPAEETAVPAVPAAPAPP 282

```

Search completed: April 4, 2003, 14:18:13
Job time : 92 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 16:26:07 ; Search time 2721 Seconds
(without alignments)
10471.021 Million cell updates/sec

Title: US-09-803-589-5

Perfect score: 979

Sequence: 1 gaattcgacagaggccgcg.....gaggatccctaccctctggc 979

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_hgt:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pi:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_srs:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_jm:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pi:*
- 26: em_ro:*
- 27: em_srs:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_hgt_hum:*
- 31: em_hgt_inv:*
- 32: em_hgt_other:*
- 33: em_hgt_mus:*
- 34: em_hgt_pln:*
- 35: em_hgt_rod:*
- 36: em_hgt_mam:*
- 37: em_hgt_vrt:*
- 38: em_sy:*
- 39: em_hgt_hum:*
- 40: em_hgt_mus:*
- 41: em_hgtg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	959.8	98.0	998	6	AX083428	Sequence
2	941.6	96.2	993	9	BC002851	BC002851 Homo sapi
3	940	96.0	969	9	AF181562	AF181562 Homo sapi
4	776.8	79.3	780	6	AX083418	Sequence
5	599	61.2	1027	10	AF181560	AF181560 Mus muscu
6	583	59.6	945	10	AF181561	AF181561 Rattus no
7	580.2	59.3	975	10	BC012263	BC012263 Mus muscu
8	502.2	51.3	1025	10	AF293356	AF293356 Mus muscu
9	474.2	48.4	113853	9	AF196971	AF196971 Homo sapi
10	343	35.0	62396	2	AC098339	AC098339 Rattus no
11	343	35.0	77876	2	AC098473	AC098473 Rattus no
12	343	35.0	111948	2	AC094563	AC094563 Rattus no
13	332.6	34.0	151694	10	AL670169	AL670169 Mouse DNA
14	308.6	31.5	177613	2	AC123014	AC123014 Rattus no
15	276.6	28.3	301	6	AX336215	AX336215 Sequence
16	276.6	28.3	301	6	AX409054	AX409054 Sequence
17	143.6	14.7	77876	2	AC098473	AC098473 Rattus no
18	122.4	12.5	177613	2	AC123014	AC123014 Rattus no
19	99	10.1	210215	2	AC097157	AC097157 Rattus no
20	98.2	10.0	206773	2	AC126296	AC126296 Rattus no
21	93.2	9.5	74721	2	AC096161	AC096161 Rattus no
22	92.2	9.4	991	11	PM12H12B	AL684455 Penicilli
23	92	9.4	136551	2	AC048354	AC048354 Homo sapi
24	91.4	9.3	174410	2	AC128158	AC128158 Rattus no
25	91	9.3	184402	2	AC127041	AC127041 Rattus no
26	90.6	9.3	74721	2	AC096161	AC096161 Rattus no
27	90.2	9.2	110958	2	AC118521	AC118521 Rattus no
28	90	9.2	1393	11	PM11H12G	AL684464 Penicilli
29	89.8	9.2	151730	2	AC125640	AC125640 Rattus no
30	88.6	9.1	151996	2	AC126885	AC126885 Rattus no
31	88.6	9.1	167527	2	AC121745	AC121745 Rattus no
32	88.2	9.0	253464	2	AC11916	AC11916 Rattus no
33	88	9.0	93419	2	AC121747	AC121747 Rattus no
34	87.8	9.0	135733	2	AC128114	AC128114 Rattus no
35	87.8	9.0	158897	2	AC121439	AC121439 Rattus no
36	87.6	8.9	10444	14	PVI422133	AJ422133 Pseudorab
37	87.4	8.9	85268	2	AC022648	AC022648 Homo sapi
38	87.4	8.9	158897	2	AC121439	AC121439 Rattus no
39	86.8	8.9	1094	11	PM7G11B	AL685196 Penicilli
40	86.6	8.8	141905	2	AC110101	AC110101 Rattus no
41	86.6	8.8	147622	2	AC110437	AC110437 Rattus no
42	86.4	8.8	61617	2	AC120284	AC120284 Rattus no
43	85	8.7	956	11	PM2D12B	AL684743 Penicilli
44	85	8.7	35487	2	AC110453	AC110453 Rattus no
45	85	8.7	52683	2	AC131481	AC131481 Rattus no

ALIGNMENTS

RESULT 1	AX083428	Sequence 120 from Patent WO0112660.	998 bp	DNA	linear	PAT 28-FEB-2001
LOCUS	AX083428	Sequence 120 from Patent WO0112660.				
DEFINITION	AX083428	Sequence 120 from Patent WO0112660.				
ACCESSION	AX083428	Sequence 120 from Patent WO0112660.				
VERSION	AX083428.1	GI:13185274				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1 (bases 1 to 998)					
AUTHORS	Kato, S. and Kimura, T.					
TITLE	Human proteins having hydrophobic domains and dnas encoding these proteins					

JOURNAL Patent: WO 0112660-A 120 22-FEB-2001;
 FEATURES SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
 source Location/Qualifiers
 1. 998
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 CDS 50. .832
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC33311.1"
 /db_xref="GI:13185275"
 /translation="MAGSPILGPPRAGGVGLVLLLLGLFRPPALCARPVKPERGLS
 AASPPLAETGAPFRFRSVPRGAAVQELARALAHLEAERQERARAQEAQOQ
 ARYLQAQVYDGPSPADLDDDDPDAPQAQALLRALRDLPALAAQLVPAVPV
 AALRPPLVYDGPADPADEAGDETPOVDPELLRYLLGRILAGSADSGEVAAPRL
 RRAADHDVGSLELPEGVIGALLARVKLETPAQVPARRLLP"
 BASE COUNT 120 a 394 c 347 g 137 t
 ORIGIN
 Query Match 98.0%; Score 959.8; DB 6; Length 998;
 Best Local Similarity 99.3%; Pred. No. 2.9e-104;
 Matches 958; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Qy 15 GCCAGCCAGTCCGCCGYMCMRRGCCGCTCGCTGGGCGACATGCGGGGTCCCGCT 74
 Db 7 GCCAGCCAGTCCGCCGTCGCGAGCCCGGCTCGCTGGGCGACATGCGGGGTCCCGCT 66
 Qy 75 GCTCTGGGGCCGGCGCGCGCTCGGCGCTTTTGGTGTCTGCTGCTGCTGCTGCGCCCTTT 134
 Db 67 GCTCTGGGGCCGGCGCGCGCGCTTTTGGTGTCTGCTGCTGCTGCTGCTGCGCCCTTT 126
 Qy 135 TCGGCGCGCCCCCGCTCTCGCGCGCGCGCGTAAAGAGCGCCCGCGCCCTAAAGCGCAGC 194
 Db 127 TCGGCGCGCCCCCGCTCTCGCGCGCGCGGTAAAGAGCGCCCGCGCCCTAAAGCGCAGC 186
 Qy 195 GTCTCGGCCCTTGCTGAGACTGCGGCTCTCTCGCGCGCTTCGGCGGTCTGAGTGCCTCCGAGG 254
 Db 187 GTCTCGGCCCTTGCTGAGACTGCGGCTCTCTCGCGCGCTTCGGCGGTCTGAGTGCCTCCGAGG 246
 Qy 255 TGAGCGCGGGGCGGTGTCAGGAGCTTGGCGCGCGCGCTGGCGCATCTGCTGGAGGCGCGA 314
 Db 247 TGAGCGCGGGGCGGTGTCAGGAGCTTGGCGCGCGCGCTGGCGCATCTGCTGGAGGCGCGA 306
 Qy 315 ACGTGAGGAGCGGCGGGCGCGAGGCGCGAGGCGTGAAGATCAGCAGGCGCGCGCTCT 374
 Db 307 ACGTCAGGAGCGGCGGGCGCGAGGCGCAGGAGCTGAGATCAGCAGGCGCGCGCTCT 366
 Qy 375 GGCGCAGTGTGCGGCTGTGGGCGGCGCCCGCACTGTATCCGGCTCTGGGCTTGGGA 434
 Db 367 GGCGCAGTGTGCGGCTGTGGGCGGCGCCCGCACTGTATCCGGCTCTGGGCTTGGGA 426
 Qy 435 CGACGACCCGACGCGCTTCAGCGGAGCTGCTGCTGCGGCTCTGCTCGCGCGCGCGCTTGA 494
 Db 427 CGACGACCCGACGCGCTTCAGCGGAGCTGCTGCTGCGGCTCTGCTCGCGCGCGCGCTTGA 486
 Qy 495 CCTTCGCGCCCTAGCAGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGCGCTCCGACC 554
 Db 487 CCTTCGCGCCCTAGCAGCCAGCTTGTCCCGCGCGCGCTCCCGCGCGCGCTCCGACC 546
 Qy 555 CCGGCGCCCGGTCTACGACGACGCGCCCGCGGGCCCGGATGCTGAGGAGGCGAGCGCAGCA 614
 Db 547 CCGGCGCCCGGTCTACGACGACGCGCCCGCGGGCCCGGATGCTGAGGAGGCGAGCGCAGCA 606
 Qy 615 GACACCGAGGTGACCCCGAGCTGTGTGAGTACTTCTGGGACGGATTCCTGGCGGAAG 674
 Db 607 GACACCGAGGTGACCCCGAGCTGTGTGAGTACTTCTGGGACGGATTCCTGGCGGAAG 666
 Qy 675 CGCGGACTCGAGGGGTGGCAGCCCGCGCGCGCTCCGCGCTGCGCGCGCGCACACGATGT 734
 Db 667 CGCGGACTCGAGGGGTGGCAGCCCGCGCGCGCTCCGCGCTGCGCGCGCGCACACGATGT 726
 Qy 735 GGCGTCTGAGTCTGCCCTCGAGGCGGTGCTGGGGCGCTGCTGCGTGTGAACGCGCTAGA 794

```
CDS      30..812
/codon_start=1
/product="granin-like neuroendocrine peptide precursor"
/protein_id="AAH02851.1"
/db_xref="GI:12804001"
/translation="MAGSPLLWGPRAGGVLLVLLGLFRPPPALCARVKEPRGLS
AASPLAETGAPRRFRSVPGEAGAVOELARALAHLLLEAEQERARAEADQO
ARVLAQLLRWGNPRNSDPALGLDDPDAPAAQLARALLRLDPAALAAQLVPAPVP
AALRRPFPVDDPGADAEAGDETDVDPELLRLLGLRILAGSADSEGVAAAPRL
RRAADHDVGSSELPPPEGVLGALLRKRLTTPAPQVPAARRLLPP"
BASE COUNT 136 a 381 c 343 g 133 t
ORIGIN
Query Match 96.2%; Score 941.6; DB 9; Length 993;
Best Local Similarity 99.7%; Pred. No. 4e-102;
Matches 941; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 36 RRGCCCGCTCGCTGGGGCAGATGGCGGGTGCCTGCTGCGGGCGCGGGCGG 95
Db 8 GGGCCCGCTCGCTGGGGCAGATGGCGGGTGCCTGCTGCGGGCGCGGGCGG 67
QY 96 GGGCGTCGGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 155
Db 68 GGGCGTCGGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 127
QY 156 CGCGCGCGCGTAAAGAGCGCCCGCGCTAAAGCGCAGCTGCTCGCGCTGAG 215
Db 128 CGCGCGCGCGTAAAGAGCGCCCGCGCTAAAGCGCAGCTGCTCGCGCTGAG 187
QY 216 TGGCGCTCCTCGCGCTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 275
Db 188 TGGCGCTCCTCGCGCTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 247
QY 276 GAGCTGCGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 335
Db 248 GAGCTGCGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307
QY 336 CGAGCGCAGGAGCTGAGATGAGCAGCGCGCTGCTGCGCGAGCTGCTGCGCT 395
Db 308 CGAGCGCAGGAGCTGAGATGAGCAGCGCGCTGCTGCGCGAGCTGCTGCGCT 367
QY 396 GGGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 455
Db 368 GGGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 427
QY 456 AGCGAGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 515
Db 428 AGCGAGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487
QY 516 GCTTGTCCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 575
Db 488 GCTTGTCCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 547
QY 576 CGGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 635
Db 548 CGGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 607
QY 636 GCTGTTGAGTACTTGTGCGAGCGATTCTTGGGGAAGCGCGAGCTCCGAGGG 695
Db 608 GCTGTTGAGTACTTGTGCGAGCGATTCTTGGGGAAGCGCGAGCTCCGAGGG 667
QY 696 AGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 755
Db 668 AGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727
QY 756 GGGCGTCTGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 815
Db 728 GGGCGTCTGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 787
QY 816 TGCACGCGCGCTTGTGCGACCTCTGAGCACTGCGCGGATCCCGTGCACCTT 875
Db 788 TGCACGCGCGCTTGTGCGACCTCTGAGCACTGCGCGGATCCCGTGCACCTT 847
```

```
QY 876 AGTGCCCGCCGCTATCCCGCAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTAC 935
Db 848 AGTGCCCGCCGCTATCCCGCAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTAC 907
QY 936 CCGCGCCAGCAGCGCTCTCACCCGAGATCCCTTACCCCTGGC 979
Db 908 CCGCGCCAGCAGCGCTCTCACCCGAGATCCCTTACCCCTGGC 951
RESULT 3
AF181562 969 bp mRNA linear PRI 27-JAN-2000
LOCUS Homo sapiens proSAAS mRNA, complete cds.
DEFINITION AF181562
ACCESSION AF181562
VERSION AF181562.1 GI:6653212
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 969)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devi,L.A. and Douglass,J.
TITLE Identification and characterization of proSAAS, a granin-like
neuroendocrine peptide precursor that inhibits prohormone
processing
JOURNAL J. Neurosci. 20 (2), 639-648 (2000)
MEDLINE 20098938
PUBMED 10632593
REFERENCE 2 (bases 1 to 969)
AUTHORS Fricker,L., McKinzie,A.A., Sun,J., Curran,E., Qian,Y., Yan,L.,
Patterson,S.D., Courchesne,P.L., Richards,B., Levin,N., Mzhavia,N.,
Devi,L.A. and Douglass,J.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1999) Department of Molecular Pharmacology,
Albert Einstein College of Medicine, 1300 Morris Park Avenue,
Bronx, NY 10461, USA
FEATURES
source Location/Qualifiers
1..969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="FLHSAAS"
1..26
27..809
/function="Inhibits prohormone processing"
/notes="granin-like neuroendocrine peptide precursor"
/codon_start=1
/product="proSAAS"
/protein_id="AAF22643.1"
/db_xref="GI:6653213"
/translation="MAGSPLLWGPRAGGVLLVLLGLFRPPPALCARVKEPRGLS
AASPLAETGAPRRFRSVPGEAGAVOELARALAHLLLEAEQERARAEADQO
ARVLAQLLRWGNPRNSDPALGLDDPDAPAAQLARALLRLDPAALAAQLVPAPVP
AALRRPFPVDDPGADAEAGDETDVDPELLRLLGLRILAGSADSEGVAAAPRL
RRAADHDVGSSELPPPEGVLGALLRKRLTTPAPQVPAARRLLPP"
sig_peptide 27..125
3'UTR 810..969
BASE COUNT 117 a 379 c 338 g 135 t
ORIGIN
Query Match 96.0%; Score 940; DB 9; Length 969;
Best Local Similarity 99.2%; Pred. No. 6.2e-102;
Matches 940; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 32 YMCRRRCGCCGCTGCTGGGGCAGCATGGCGGCTGCCGCTCTGGGGCGCGGG 91
Db 1 TCCGGAGCAGGCTGCTGGGGCAGCATGGCGGCTGCCGCTCTGGGGCGCGGG 60
QY 92 CCGGGCGCTGGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 151
Db 61 CCGGGCGCTGGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
```



```
source 1. .945
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="FLRSAAS"
1. .21
22. .804
/function="inhibits prohormone processing"
/notes="granin-like neuroendocrine peptide precursor"
/codon_start=1
/product="proSAAS"
/protein_id="AAE22642.1"
/db_xref="GI:6653211"
/translaton="MAGSPLLCPRAGVGLVLLGLLRLPTLSARPVKPRSL
AASPLAETPLRLRAVRGEAGAVELARALAHLEAERARARAOAEADQ
ARVLAQLRWGSPRASDPPLADDDPAPAAQALRALRLDPAALAAQVLPAP
AALRRPPVDDGPTGPDVEDAADPTDVPDPELLRLLGLILTSSEPEAPAPRL
BRAVDODLGPVEVPENVLGALLRVKRLSNSPQAPARLLPP"
22. .120
sig_peptide
3'UTR 805. .945
polyA_signal 921. .926
polyA_site 939
BASE COUNT 134 a 355 c 307 g 149 t
ORIGIN
Query Match 59.68; Score 583; DB 10; Length 945;
Best Local Similarity 77.5%; Pred. No. 6e-60;
Matches 714; Conservative 1; Mismatches 205; Indels 1; Gaps 1;
QY 37 RGCCTGCTGCTGGGAGCATGCGGGTCTGCGCTGCTGCGGGGCGCGGGCGGG 96
Db :|||||
QY 1 AGCCGGCTGCTGGGAGCATGCGGGTCTGCGCTGCTGCGGGGCGCGGGCGGG 60
Db :|||||
QY 97 GCGCTGCGGCTTTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 156
Db :|||||
QY 61 GCGTAGGCTTTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTA 120
Db :|||||
QY 157 GCGGGCGGTAAGAGAGCCCGGGCTTAAGCGAGGCTTCGCCCTTGCTGGCTGAGACT 216
Db :|||||
QY 121 GCGAGGCGGTGAAGAACCCGAGTCTGAGCGCAGCATCCGGCGCTTGGCTGAGACG 180
Db :|||||
QY 217 GCGCTCTCTCCGCTTCCGCGCTGAGTCCCGAGTGGAGGCGGGGCGGTGCGAG 276
Db :|||||
QY 181 AGCACTCCCTTCCGCTGCTGCGGGCTGCGCGAGGAGGCGGGTGGCGGTGCGAG 240
Db :|||||
QY 277 GAGTGGCGGGGCGCTGGCGCATCTGCTGAGGCGCGAAGCTCAGGAGCGGGCGGGGCC 336
Db :|||||
QY 241 GAGTGGCGGGGCGCTGGCGCATCTGCTGAGGCGCGAGAGACAGAGCGCGCGTGTCT 300
Db :|||||
QY 337 GAGCGCAGGAGGCTGAGGATCAGCAGCGCGGCTGCTGCGGCGAGCTGCTGCGGCTG 396
Db :|||||
QY 301 GAGCGCAGGAGGCGGAGGATCAGCAGGCGAGGCTGCTGCGCAGAGCTGCTGCGGCG 360
Db :|||||
QY 397 GCGCGCCCGCCCAACTCTGATCCGGCTTGGGCTTGGAGCAGCAGCCCGCGCGCTGCA 456
Db :|||||
QY 361 GCGTCCCGGCTGCTGCGACCCCGCTTGGCCCGGATGACGACCGCGAGCTCCGCT 420
Db :|||||
QY 457 GCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 516
Db :|||||
QY 421 GCACAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db :|||||
QY 517 CTTCCTCCCGCGCGCTCCCGCGCGGCTCCGACCCCGCGCGCGCGCGCTACGACGAC 576
Db :|||||
QY 481 CTTCCTCCCGCGCGCTCCCGCGCGGCTCCGACCCCGCGCGCGCGCTATGACGAT 540
Db :|||||
QY 577 GCGCGCGCGCGCGGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 636
Db :|||||
QY 541 GCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
Db :|||||
QY 637 CTGTTGAGGTACTTGTGGGAGGATTTTGGGAGAGCGCGACTCCGAGGGGGTGGCA 696
Db :|||||
QY 601 CTGCTGAGGTACTTGTGAGGCGGATCTCACCGAGTTCGGAGCGCAGGCTGCTCCG 660
Db :|||||
QY 697 GCGCGCGCGCGCTCCCGCGCGGCGGACACAGATGTGGGCTGAGCTGCGCCCTGAG 756
Db :|||||
Db 661 GCCCGCGCGCTCCCGCGGAGCTGTGACGAGGATTTGGTCCGAGGTGCTCTCTGAG 720
QY 757 GCGGTGCGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
Db 721 AAGTACTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 817 GCAGCGCGCTTTCGCCACCTGACACTGCGCGGATCCCGTGCACCTGGGACCCAGAA 876
Db 781 GCAGCGCGCTTTCGCCACCTGACACTGCGCGGATCCCGTGCACCTGGGACCCAGAA 839
QY 877 GTGCGCGCGCATCCCGCCACGAGACTGCTCCCGCGCAGCAGCTCCAGAGCAACTTACC 936
Db 840 CGCGCGCAGCAACCTGACTCCCTGCCAGCAGCTCCAGCGGTCTTACCCGCGCAACCTC 899
QY 937 CCGCGCAGCCAGCTCTCTAC 957
Db 900 CCATCCCGTGGAGCGCTCCCTC 920
RESULT 7
BC012263 975 bp mRNA linear ROD 07-AUG-2002
LOCUS BC012263
DEFINITION Mus musculus, Similar to granin-like neuroendocrine peptide precursor, clone MGC:19107 IMAGE:4207854, mRNA, complete cds.
ACCESSION BC012263
VERSION BC012263.1 GI:15126669
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.
Direct Submission
Submitted (06-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAC Plate: 24 Row: p Column: 24.
FEATURES
Location/Qualifiers
source 1. .975
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="MGC:19107 IMAGE:4207854"
/tissue_type="Colon, normal. 5 month old male mouse."
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
20. .796
/codon_start=1
/product="Similar to granin-like neuroendocrine peptide precursor"
/protein_id="AAH12263.1"
/db_xref="GI:15126670"
CDS
```

```
/translation="MAGSPLLCGPRAGVGILVLLGLLRLPPTLSARPVKEPSLS
AASAPLVETSTPLRLRAVPGEAAGAQBELARALAHLEERQERARAEQAEQDQ
ARVLQALLRWGSPRASDPPLAPDDPDAPAAQALRALRLARLDPAAALAAQVLPAPAA
APRPFPYDGDPTGPDVEDAGDETPDVPDPELLRYLLGRILTGSSEPAAPAPRLRR
SYVDPLGPEVPENIVGALLRVKLENFSPQAPARRLLPP"
BASE COUNT      173 a 353 c 304 g 145 t
ORIGIN

Query Match      59.3%; Score 580.2; DB 10; Length 975;
Best Local Similarity 80.0%; Pred. No. 1.3e-59;
Matches 726; Conservative 0; Mismatches 163; Indels 18; Gaps 3;

QY 39 CCGCGCTCGTGGGCGAGCATGGCGGCTCGCGCTGCTTGGGGCCCGCGGCGGGG 98
DB 1 CCGGGCTCGTGGGCGAGCATGGCGGGTGGCGCTGCTTGGGGCCCGCGGCGGGG 60
QY 99 CGTGGCCCTTTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 158
DB 61 CGTGGCCATTTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 159 GCGGCGGTTAAGGACCCCGGCGCTAAGCGAGCGTCTCGCCCTTGGCTGAGACTGG 218
DB 121 GAGGCGCGTGAAGACCCCGGAGTGAAGCGAGCATCGCGCCCTTGGTTGAGACGAG 180
QY 219 CGCTCTCGCGCTTCCGGCGGTCACTGCGCGAGGTGAGGCGGCGGCGGTGCAGGA 278
DB 181 CACTCCCTCGGCTTGGCGGCGTGGCGGAGAGGCGGCGGTGCGGTGCAGGA 240
QY 279 GCTGGCGCGGCGCTGGCGCATCTGCTGAGGCGCAACGTCAAGAGCGGCGGCGCCGA 338
DB 241 GCTGGCGGCGGCTGGCGACCTGCTGGAGCGCGAGACAGGACGCGCGCTGCTGA 300
QY 339 GGGCGAGGAGGTGAGGATCAGCAGCGCGGCTGCTGGCGGAGCTGCTGGGCTTGGGG 398
DB 301 GGGCGAGGAGGTGAGGATCAGCAGCGCGGCTGCTGGCGGAGCTGCTGGCGCTGGG 360
QY 399 CGCGCCCGCAACTGATCCGCTGCTGGGCTTGGAGCAGACCGCGAGCGGCTGCAGC 458
DB 361 CTCTCGGCTGCTCGGACCGCGCTTGGCCCGGAGATGACCGGAGCTCCAGCTGC 420
QY 459 GCAGCTCGTGGCGCTGCTCGCGCGCGCTTGACCTCGCGCGCTAGCAGCCAGCT 518
DB 421 ACAGCTCGCGCGTCTGCTCGGAGCTCGCTAGACCGCGCGCGCTGCGACCCAACT 480
QY 519 TGTCCCGCGCGCTGCCCGCGGCGCTCGACACCGCGCGCGCGCGCTTACAGCACGCG 578
DB 481 TGTCCCGCG-----CCCTGCGGCTGCGCGGAGACCGCGCGCGCTATGATGATG 534
QY 579 CCGCGCGCGCGGATGCTGAGGAGCGAGCGAGACACCGGAGCTGAGCCCGAGCT 638
DB 535 CCGGCTGCGCGGAGCTGAGGATCGCGGAGAGAGCTCTGAGCTGGACCTGAGCT 594
QY 639 GTTGAAGTACTTGTGCGGAGCGGATTTTGGGGAAGCGGGAATTCGAGGGGTGCGAGC 698
DB 595 GCTGAGGTACTTGTAGGCGGATCTCACCAGGATTCGAGGACGAGGCTGCTCTGC 654
QY 699 CCGCGCGCGCTCGCGGCTGCGCGACGAGTGTGGGCTGTGAGCTGCGCCCTGAGGG 758
DB 655 CCGCGCGCGCTCGCGGATCTGTGACCAAGGATTTGGGTTCGCGAGGTGCGCCCTGAGAA 714
QY 759 CFTGCTGGGGGCGCTGCTGCTGTGAACGCTAGAGACCGCGCGCGCGCTGCGCTGC 818
DB 715 CGTACTGGGGGCTGCTGCTGAGCGCTGAACGCTGAGAGACCGCTCGCGCGCGCGG 774
QY 819 AGCGCGCTCTTGGCAACCTGAGCACTGCGCGGATCCCGTGCACCTGGGACCCAGG 878
DB 775 AGCGCGCTCTTGGCTTCTGAGCGCTG-CTGATCTTGCACGCGCTTGAACCCAGGAGC 833
QY 879 GCGCGCGCATCCCGCGCACGAGACTGTCTCCCGCGCGAGCAGCTTCCAGACAACTTACCC 938
DB 834 GCGCGCGCAACCTG-----ACTCCCTGCGAGCAGCTCCCAAGGCTGCTTACCC 882
QY 939 GCGCAGC 945
```

```
|||||
Db 883 AGCAACC 889

RESULT 8
AF293356
LOCUS      AF293356
DEFINITION Mus musculus IA4 mRNA, complete cds.
ACCESSION AF293356
VERSION    AF293356.1 GI:9954415
KEYWORDS
SOURCE     Mus musculus.
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1025)
Donadel,G., Marinos,N., Desilva,M.G., Lu,J., Notkins,A.L. and
Lan,M.S.
Molecular cloning and characterization of a highly basic protein,
IA-4, expressed in pancreatic islets and brain
Neuroendocrinology 67 (3), 190-196 (1998)
9630436
REFERENCE 2 (bases 1 to 1025)
AUTHORS   Donadel,G., Notkins,A.L. and Lan,M.S.
DIRECT SUBMISSION
TITLE      Submitted (03-AUG-2000) Laboratory of Oral Medicine, National
Institute of Dental Research, 9000 Rockville Pike, Building 30 Room
124, Bethesda, MD 20892, USA
FEATURES
source     1..1025
            Location/Qualifiers
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /cell_line="insulinoma b-TC-1"
            94..657
            /note="highly basic protein; highly expressed in
            pancreatic islets and neuronal cells"
            /codon_start=1
            /product="IA4"
            /protein_id="GI:9954416"
            /db_xref="GI:9954416"
            /translation="MAGSPLLCGPRAGVGILVLLGLLRLPPTLSARPVKEPSLS
            ARSAPLVETSTPLRLRAVPGEAAGAVPGAGCAGAPAGGDRNARVLRRLRLIS
            RMRWSRSCAFGALRVTRPLAPDDPDAPAAQALRALRLARLDPAAALAAQVLPAPA
            PAPRPFPYDGDPTGPDVEDAGDETPDVPDPELLRYLLGRILTGSSEPAAPARRLLPP"
BASE COUNT 156 a 382 c 330 g 157 t
ORIGIN

Query Match      51.3%; Score 502.2; DB 10; Length 1025;
Best Local Similarity 78.0%; Pred. No. 1.9e-50;
Matches 732; Conservative 5; Mismatches 173; Indels 29; Gaps 10;

QY 15 GCAGCAGCTCCGCCGCGMCGRRGCCGCTCGCTGGGCGAGCATGGCGGGTCCGCGT 74
DB 51 GACAGCAGCTCCGCCGCTCCGAGCGCGGCTCGTTGGGCGAGCATGGCGGGTCCGCGT 110
QY 75 GCTCTGGGCGCGCGGCGGCGCTCGCGCTTTTGGTGTCTGCTGCTGCTGCGGCTGTT 134
DB 111 GCTCTGGGCGCGCGGCGGCGCTCGGCATTTTGGTGTCTGCTGCTTGGGCGCTTCT 170
QY 135 TCGGGCGCGCGCGCGCTCTGCGCGCGCGGTAAAGAGACCGCGCGCTAAGCGCAGC 194
DB 171 GAGGCTGCCCGCCACCTGTGACGAGAGCCCGGTGAAGAGAGCCCGGAGTGTGAGCAGC 230
QY 195 GTCTCCGCGCTTGGCTGAGACTGGCGCTCCTCGCGCTTCCGCGGCTCAGTGCCTGAGG 254
DB 231 ATCCGCGCGCTTGGTTGAGACGAGCACTCCCTCCGCTTGGCTGCGGCGCTGCCCGCAGG 290
QY 255 TGAGGCGCGGCGGCGGTG-CAGGAGCTGGCG---GGCGCTGGCGCATCTGCTGGAGG 310
DB 291 AGAGCGCGGCGGTGCGGTGCCAGAGAGTGGCGCGGCGGCGCTGGCGACCTGCTGGAGG 350
QY 311 CCGAACGCTCAGGAGCGGCGCGGCGCGAGCGCTGAGGATCAGAGCGCGCGCG 370
```

Db 351 CCGA--GACAGGACCGCGCTGTAGGCGCAGGAGCTGAGGATCAGAGCGCATG 408
Qy 371 TCCTGGCGACGCTGTGGCGCTCTGGGCGCCCGCCGCAACTCTGATCCG-GCTCTGGCG 429
Db 409 TCCTGGCGACGCTGTGGCGCCCTGGGCGCTCTGGGCGCTCGGACCGCGCTTGGCC 468
Qy 430 TTGGACGACGACCCGCGGCTGACAGCGCTGCGAGCTGCGGCTCTGCTCCGCGCGCGC 489
Db 469 CCGGACGATGACCGGCGGCTGACAGCTGACAGCTGCGCGCTGCTGCTCCGAGCTCGC 528
Qy 490 CTTGACCTGCGCGCTGACGACCGCTGCTGCGCGCGCGCGCTGCGCGCGCGCGCTC 549
Db 529 CTAGACCGCGCGCGCTGCGCGCGCGCAACTTGTCCCGC-----CCCTGCGCTGCGCGC 582
Qy 550 CGACCCGCGCGCGCGCTGACAGCA--CGGCGCGCGCGCGCGGATGCTGAGGAGCGAGG 608
Db 583 CGACCCGCGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 642
Qy 609 CGACGACGACCCGCGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 668
Db 643 CGACGAGACTGGTGAAGCT-GACCCCTGAGCTGCTGAGGTAAGTCTGCTAGGCGGATCCCTCAC 701
Qy 669 GGGAGCGCGGACCTCCGAGGGGCTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 728
Db 702 CGGAAGTTGCGGAGCGCAGAGGCTGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGC 761
Qy 729 CGATCTGGGCTCTGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 786
Db 762 GGAATTTGGC 821
Qy 787 GCGCTAGAGACCCGC 846
Db 822 CGCCTGGAGAACCCCTGC 881
Qy 847 CCCGATCCCGTGCACCTGGGACCGCAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGC 906
Db 882 -CTGATCTCTGACCGC 929
Qy 907 TCCCGCGCAGCAGCTCCGAGCACTTACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 945
Db 930 TCCCTGCCAGCAGCTCCGAGCGCTTACCCCGCGCGCGCGCGCGCGCGCGCGCGCGC 968

RESULT 9
AF196971/c
LOCUS
DEFINITION
Homo sapiens GATA-binding protein 1 and histone deacetylase-like protein genes, complete cds; CRAS pseudogene, complete sequence; and protein translocase gene, partial cds.
ACCESSION
VERSION AF196971
KEYWORDS HTG.
SOURCE AF196971.1 GI:6289080
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 113853)
AUTHORS Blechschmidt,K., Nyakatura,G., Strom,T.M., Drescher,B., Menzel,U., Weindl,A. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (14-Oct-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT 1-36552: cosmid J1045; 11106-52106: cosmid E2237; 27017-67034: cosmid L1850; 58536-92644: cosmid D1425; 80573-113853: cosmid E1239.
FEATURES
source Location/Qualifiers
1..113853
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="xp11.23"
/clone="cosmids J1045; E2237; L1850; D1425; E1239"

repeat_region 1..252
/evidence=not_experimental
/rpt_family="AluSx"
254..396
repeat_region
/evidence=not_experimental
/rpt_family="L1M4"
397..704
repeat_region
/evidence=not_experimental
/rpt_family="AluSg"
705..717
repeat_region
/evidence=not_experimental
/rpt_family="L1M4"
817..1083
repeat_region
/evidence=not_experimental
/rpt_family="L1MB7"
complement(1084..1371)
repeat_region
/evidence=not_experimental
/rpt_family="AluJo"
1372..1790
repeat_region
/evidence=not_experimental
/rpt_family="L1MB7"
1802..1924
repeat_region
/evidence=not_experimental
/rpt_family="L1MD2"
complement(2269..2440)
repeat_region
/evidence=not_experimental
/rpt_family="L2"
complement(2950..3059)
repeat_region
/evidence=not_experimental
/rpt_family="MIR"
3105..3235
repeat_region
/evidence=not_experimental
/rpt_family="MIR"
3268..3472
repeat_region
/evidence=not_experimental
/rpt_family="AluJo"
3504..3608
repeat_region
/note="homology = 80.00%, score = 28, counts = 3"
/evidence=not_experimental
/rpt_type=tandem
/rpt_unit-ttagcattaggaatgatttactataataaattagta
repeat_region
3613..3955
/evidence=not_experimental
/rpt_family="L1MA5"
3964..4059
repeat_region
/note="homology = 77.10%, score = 20, counts = 3"
/evidence=not_experimental
/rpt_type=tandem
/rpt_unit-aataatattagtagtattactataatcagtat
repeat_region
complement(4121..4369)
/evidence=not_experimental
/rpt_family="MIR"
4393..4463
repeat_region
/evidence=not_experimental
/rpt_family="L2"
complement(4502..4544)
repeat_region
/evidence=not_experimental
/rpt_type=tandem
/rpt_unit-tc
5649..5691
repeat_region
/evidence=not_experimental
/rpt_family="MIR"
5062..5157
/note="homology = 67.70%, score = 32, counts = 48"
/evidence=not_experimental
/rpt_type=tandem
/rpt_unit-tc
5649..5691
repeat_region
/evidence=not_experimental
/rpt_family="MIR"
6131..13747
gene
/gene="GATA-1"
mRNA
join(6131..6168,10614..10852,11367..11744,11846..11991,
12695..12820,13316..13747)
/gene="GATA-1"
/product="GATA-binding protein 1"
/note="mRNA acc. no. M30601"

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 62396)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,F., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oraguine,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Petters,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sison,I.,
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tantey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

Direct Submission
2 (bases 1 to 62396)
Worley,K.C.

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

JOURNAL

Direct Submission
3 (bases 1 to 62396)
Worley,K.C.

Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 24, 2002 this sequence version replaced gi:21953722.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TULN
Center clone name: CH230-2A16
----- Summary Statistics
Sequencing vector: plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329

Consensus quality: 23320 bases at least Q40
Consensus quality: 26055 bases at least Q30
Consensus quality: 27202 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1088: contig of 1088 bp in length
* 1089: gap of unknown length
* 1189: contig of 1144 bp in length
* 2332: gap of unknown length
* 2333: contig of 1205 bp in length
* 3637: gap of unknown length
* 3638: contig of 1147 bp in length
* 3738: gap of unknown length
* 4885: gap of unknown length
* 4885: contig of 1247 bp in length
* 6232: gap of unknown length
* 6332: contig of 1066 bp in length
* 7398: gap of unknown length
* 7498: contig of 1105 bp in length
* 8603: gap of unknown length
* 8702: contig of 1006 bp in length
* 8703: gap of unknown length
* 9709: contig of 1287 bp in length
* 9809: gap of unknown length
* 11096: contig of 1394 bp in length
* 12590: gap of unknown length
* 12690: contig of 1178 bp in length
* 13868: gap of unknown length
* 13968: contig of 1183 bp in length
* 15151: gap of unknown length
* 15251: contig of 1141 bp in length
* 16392: gap of unknown length
* 16392: contig of 1037 bp in length
* 17529: gap of unknown length
* 17629: contig of 1278 bp in length
* 18907: gap of unknown length
* 19007: contig of 1146 bp in length
* 20153: gap of unknown length
* 20253: contig of 1040 bp in length
* 21293: gap of unknown length
* 21393: contig of 1308 bp in length
* 22701: gap of unknown length
* 22801: contig of 1906 bp in length
* 24707: gap of unknown length
* 24807: contig of 1582 bp in length
* 26389: gap of unknown length
* 26489: contig of 1149 bp in length
* 27638: gap of unknown length
* 27738: contig of 1365 bp in length
* 29103: gap of unknown length
* 29203: contig of 1150 bp in length
* 30353: gap of unknown length
* 30453: contig of 1268 bp in length
* 31721: gap of unknown length
* 31821: contig of 2083 bp in length
* 33904: gap of unknown length
* 34004: contig of 1756 bp in length
* 35760: gap of unknown length
* 35860: contig of 1381 bp in length
* 37341: gap of unknown length
* 39223: contig of 1882 bp in length
* 39323: gap of unknown length
* 40924: contig of 1601 bp in length
* 41024: gap of unknown length
* 42882: contig of 1859 bp in length

```

* 42883 42982: gap of unknown length
* 42983 44647: contig of 1665 bp in length
* 44647 44747: gap of unknown length
* 44748 46546: contig of 1799 bp in length
* 46547 46646: gap of unknown length
* 46647 48461: contig of 1815 bp in length
* 48462 48561: gap of unknown length
* 48562 51223: contig of 2662 bp in length
* 51224 51343: gap of unknown length
* 51344 53891: contig of 2568 bp in length
* 53892 53991: gap of unknown length
* 53992 56351: contig of 2360 bp in length
* 56352 56451: gap of unknown length
* 56452 58948: contig of 2497 bp in length
* 58949 62396: gap of unknown length
* 59049 62396: contig of 3348 bp in length.
FEATURES
    source
    1. .62396
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-2A16"
BASE COUNT 15738 a 13739 c 13230 g 15967 t 3722 others
ORIGIN
Query Match 35.0%; Score 343; DB 2; Length 62396;
Best Local Similarity 82.3%;
Matches 394; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 170 AGGAGCGCCGCGCGCTAAGCGCAGCGCTCTCGGCCCTTGGCTGAGACTGGCGCTCCCTCGCC 229
Db 54162 AGGACCCCGCAGTCTGAGCGAGCATCCGCGCCCTTGGCTGAGAGCAGCTCCCTCC 54221
QY 230 GCTTCGCGCGCTCAGTGCCTCCCGAGGTGAGCGCGCGCGCGGTGTCAGGAGCTGGCGCGG 289
Db 54222 GCTTCGCTCGCGCGCTGCGCGAGAGAGCGCGGTGCGGTGTCAGGAGCTGGCGCGG 54281
QY 290 CGCTGGCGCATCTGCTGGAGCGCGAGCTCAGGAGCGCGCGCGCGCGCGAGGCGGAGG 349
Db 54282 CGCTGGCGCATCTGCTGGAGCGCGAGAGCAGAGCGCGCGCGGTGCTGAGGCGAGG 54341
QY 350 CTGAGGATCAGCAGCGCGCGCTCTGCGCGAGCTGCTGCGCGCTGCGCGCGCGCGCGCGCA 409
Db 54342 CGGAGGATCAGCAGCGCGGTCTGCGCAGCTGCTGCGCGCTGCGCGCTCCCGCGTG 54401
QY 410 ACTCTGATCCGCTCTGGCTTGGAGCAGCAGCCCGCGCGCGCTGCGAGCGCAGCTCGCTC 469
Db 54402 CCTCGACCGCGCTTGGCCCGGATGACGACCGCGCGCTCCCGCTGCGACAGCTCGCC 54461
QY 470 GCGCTGCTCCGCGCGCGCTTGAACCTGCGCGCTAGCAGCCGCGCTGCTCCCGCGCG 529
Db 54462 GTGCTGCTCCGAGCTGCGCTGAGACCGCGCGCGCTGCGCGCTGCTGCTCCCGCGCG 54521
QY 530 CCGTCCCGCGCGCGCGCTCCGACCGCGCGCGCGCTCTACGACGAGCGCGCGCGCGCGCC 589
Db 54522 CTGCGCTGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCTGATGAGTGGCGCGCGCGCC 54581
QY 590 CGGATCTGAGGAGGAGCGCGCAGCAGCAGCCCGCGCGCGCGCTGAGCGCTGTCAGGATAC 648
Db 54582 CAGACGTCGAGGATGCGCTGAGCAGACTCCTGATGTGGACCGCTGAGCTGCTGAGGTGC 54640

RESULT 11
AC098473
LOCUS
DEFINITION
Rattus norvegicus clone CH230-2J23, *** SEQUENCING IN PROGRESS ***,
36 unordered pieces.
AC098473
AC098473.3 GI:21953422
VERSION
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
```

REFERENCE
AUTHORSRattus.
1 (bases 1 to 77876)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T., Barbarella, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Coyle, M.D., Dathorne, S., David, R., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseghed, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, K., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Roife, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williams, A., Wlezyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

TITLE
JOURNAL2 (bases 1 to 77876)
Worley, K.C.REFERENCE
AUTHORS

Submitted (23-Oct-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

TITLE
JOURNAL

Submitted (31-Jul-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 24, 2002 this sequence version replaced gi:20976551.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUVZ
Center clone name: CH230-2J23
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 42926 bases at least Q40
Consensus quality: 47276 bases at least Q30
Consensus quality: 50269 bases at least Q20

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, L.E., Homi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, E., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, F., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Roife, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmari, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 111948)
Worley, K.C.

Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 111948)
Worley, K.C.

Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17941324.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAXL
Center clone name: CH230-4X5
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 54807 bases at least Q40
Consensus quality: 60236 bases at least Q30
Consensus quality: 64720 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.

1 1001: contig of 1001 bp in length
1002 1101: gap of unknown length
1102 2332: contig of 1231 bp in length
2333 2432: gap of unknown length
2433 3485: contig of 1053 bp in length
3486 3585: gap of unknown length
3586 4814: contig of 1229 bp in length
4815 4914: gap of unknown length
4915 6154: contig of 1240 bp in length
6155 6254: gap of unknown length
6255 7917: contig of 1663 bp in length
7918 8017: gap of unknown length
8019 9057: contig of 1040 bp in length
9058 9157: gap of unknown length
9158 10364: contig of 1207 bp in length
10365 10464: gap of unknown length
10465 11897: contig of 1433 bp in length
11898 11997: gap of unknown length
11999 13263: contig of 1266 bp in length
13264 13363: gap of unknown length
13364 14544: contig of 1181 bp in length
14545 14644: gap of unknown length
14645 16064: contig of 1420 bp in length
16065 16164: gap of unknown length
16165 17785: contig of 1621 bp in length
17786 17885: gap of unknown length
17886 19123: contig of 1238 bp in length
19124 19223: gap of unknown length
19224 20631: contig of 1408 bp in length
20632 20731: gap of unknown length
20732 22214: contig of 1483 bp in length
22215 22314: gap of unknown length
22315 23749: contig of 1435 bp in length
23750 23849: gap of unknown length
23850 25423: contig of 1574 bp in length
25424 25523: gap of unknown length
25524 26850: contig of 1327 bp in length
26851 26950: gap of unknown length
26951 28487: contig of 1537 bp in length
28488 28587: gap of unknown length
28588 29836: contig of 1249 bp in length
29837 31672: contig of 1736 bp in length
31673 31772: gap of unknown length
31773 33161: contig of 1389 bp in length
33162 33261: gap of unknown length
33262 34438: contig of 1176 bp in length
34439 36752: contig of 2215 bp in length
36753 36852: gap of unknown length
36853 37909: contig of 1057 bp in length
37910 38009: gap of unknown length
38010 39170: contig of 1161 bp in length
39171 39270: gap of unknown length
39271 41040: contig of 1770 bp in length
41041 41140: gap of unknown length
41141 42441: contig of 1301 bp in length
42442 42541: gap of unknown length
42542 43610: contig of 1069 bp in length
43611 43710: gap of unknown length
43711 45019: contig of 1309 bp in length
45020 45119: gap of unknown length
45120 46358: contig of 1239 bp in length
46359 46458: gap of unknown length
46459 48595: contig of 2137 bp in length
48596 48695: gap of unknown length
48696 50326: contig of 1631 bp in length
50327 50426: gap of unknown length
50427 51957: contig of 1531 bp in length
51958 52057: gap of unknown length
52059 53961: contig of 1904 bp in length


```

QY 763 CTGGGGCGCTGCTGCGTGTGAACGCTAGAGACCCCGGCGCCAGGTGCCTGCACGC 822
Db 241 CTGGGGCGCGTGTGCGTGTGAACGCTAGAGACCCCGGCGCCAGGTGCCTGCACGC 182
QY 823 CGCCTCTTGCCACCCCTGAGCACTGCCCCGGATCCCGTGCACCCCTGGGACCCAGAAAGTGCCC 882
Db 181 CGCCTCTTGCCACCCCTGAGCACTGCCCCGGATCCCGTGCACCCCTGGGACCCAGAAAGTGCCC 122
QY 883 CGGCCATCCGGCCACCCAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGGCC 942
Db 121 CGGCCATCCGGCCACCCAGGACTGCTCCCGCCAGCAGCTCCAGAGCAACTTACCCCGGCC 62
QY 943 AGCCAGCCCTCTCACCGGAGGATCCCTACCCCTGGC 979
Db 61 AGCCAGCCCTCTCACCGGAGGATCCCTACCCCTGGC 25

```

Search completed: April 12, 2003, 18:13:55
Job time : 3115 secs

